

" AREN11.US9.DIV
" Serial No. 09/875,076

November 17, 2006 Response
to May 19, 2006 Action

APPENDIX 3

Probst et al., (1992) DNA Cell Biol.

REVIEW ARTICLE

Sequence Alignment of the G-Protein Coupled Receptor Superfamily

WILLIAM C. PROBST,*† LENORE A. SNYDER,*§ DAVID I. SCHUSTER,‡ JÜRGEN BROSIUS,*
and STUART C. SEALFON*,†

ABSTRACT

The multitude of G-protein coupled receptor (GPR) superfamily cDNAs recently isolated has exceeded the number of receptor subtypes anticipated by pharmacological studies. Analysis of the sequence similarities and unique features of the members of this family is valuable for designing strategies to isolate related cDNAs, for developing hypotheses concerning substrate-ligand and receptor-effector interactions, and for understanding the evolution of these genes. We have compiled and aligned the 74 unique amino acid sequences published to date and review the present understanding of the structural motifs contributing to ligand binding and G-protein coupling.

INTRODUCTION

THE CLONING of a great number of receptors and channels has revealed that many of these critical membrane proteins can be grouped into gene superfamilies based on sequence and structural similarities. One of these superfamilies comprises the G-protein coupled receptors (GPRs). Although the signal transduction mechanism is not known for all members of the gene family, in most cases receptor stimulation induces activation of a guanine nucleotide binding protein or G-protein. In 1982 the complete protein sequence of the visual pigment bovine rhodopsin was determined (Ovchinnikov *et al.*, 1982). Its predicted structure, containing an extracellular amino terminus and seven hydrophobic membrane spanning α -helices (Hargrave *et al.*, 1983), was remarkably similar to that previously identified by electron diffraction and sequence analysis for bacteriorhodopsin (Unwin and Henderson, 1975; Engelman *et al.*, 1980). The subsequent molecular cloning of four human opsins (Nathans and Hogness, 1984; Nathans *et al.*, 1986) and the hamster β -adrenergic receptor (Dixon *et al.*, 1986) again revealed these structural features that have become the hallmark of this gene family (Fig. 1).

The number of GPRs that have been cloned is increasing rapidly; at present 74 distinct GPR sequences have been published. GPR cloning has led to the stable high-level expression of these receptor subtypes in mammalian cell lines, a preparation that has greatly aided the pharmacological characterization of these receptors. Molecular biological alteration of receptor sequences and expression in cell lines has provided much of our knowledge concerning the functional role of particular receptor regions and residues.

We have aligned all the available amino acid sequences of the members of this family (Fig. 2). This compilation should prove useful for designing cloning strategies for other GPRs. Indeed, many GPRs, among them the dopamine receptors (Bunzow *et al.*, 1988; Dearry *et al.*, 1990), the adenosine receptors (Libert *et al.*, 1989b), and the cannabinoid receptor (Matsuda *et al.*, 1990), have been cloned *via* approaches relying on sequence similarity. In addition, this sequence alignment may facilitate the formulation of hypotheses concerning the role of certain protein sequences in determining ligand binding, regulation, and G-protein specificity of the receptors. Comparison of the structure of the genes for these receptors can provide insight into the evolution of this gene family.

*Fishberg Research Center for Neurobiology, †Department of Neurology, Mount Sinai School of Medicine, New York, NY 10029.

‡New York University, Department of Chemistry, New York, NY 10003.

§The Graduate School and University Center of the City University of New York, Ph.D. Program in Biology, New York, NY 10036-8099.

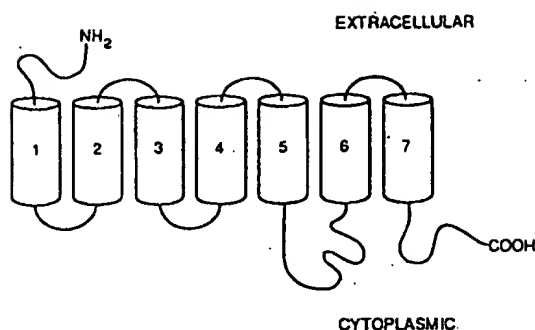


FIG. 1. The topography of G-protein linked receptors. Cylinders represent transmembrane α -helices. Extracellular and cytoplasmic sides of the plasma membrane are indicated.

The sequences were aligned manually, relying on invariant residues and published computer-generated sequence alignments. Several of the sequences, such as the FC5R receptor, have not yet been proven to represent GPRs. They are included in the alignment, however, because their sequences identify them as members of this superfamily. If sequences for the same receptor subtype of more than one species have been published, we have included only the sequence of the highest species. The sequences are organized into subgroups based on ligand type, *i.e.*, muscarinic receptors, catecholamine receptors, *etc.*

GENERAL STRUCTURAL FEATURES

All of the proteins are single polypeptide chains. The shortest sequence represents the rat *mas* oncogene (324 amino acids) and the longest sequence represents the human thyroid-stimulating hormone receptor (744 amino acids). The predicted protein structures contain seven stretches of 20–30 hydrophobic amino acids which are believed to form membrane-spanning α -helices. These helices are referred to as transmembrane domains 1–7 (TM 1–TM 7). This predicted structure, based on hydropathy analysis, has been supported by electron diffraction analysis for bacteriorhodopsin (Henderson *et al.*, 1990) and proteolytic cleavage studies for rhodopsin and the β_1 -adrenergic receptor (Hargrave *et al.*, 1982; Dohlman *et al.*, 1988). The proteins have extracellular amino termini and cytoplasmic carboxyl termini.

The areas of greatest homology among the GPRs are in the seven transmembrane regions. Some residues are found in virtually all GPRs and may mediate the tertiary structure required for functional activity (Hulme *et al.*, 1990; Hibert *et al.*, 1991). Particularly well conserved are several proline residues in TM 4, 5, 6, and 7. These residues most likely introduce kinks in the α -helices and may be important in the formation of the binding pocket (Applebury and Hargrave, 1986; Findlay and Eliopoulos, 1990; Dahl *et al.*, 1991; Hibert *et al.*, 1991). Other well-conserved residues include a glycine, an asparagine, and a valine in TM 1; a leucine, two alanines, and an aspartate in TM 2; an isoleucine in TM 3; a tryptophan in TM 4; a phenylalanine

and a tryptophan in TM 6; and an asparagine and a tyrosine in TM 7 (see Fig. 3). Certain conserved residues are replaced in particular subfamilies. For example, the TM 6 conserved tryptophan is replaced by methionine in the glycoprotein hormone receptors (Fig. 2).

Most GPRs have single conserved cysteine residues in each of the first two extracellular loops that are believed to form a disulfide bond that stabilizes the functional protein structure (see Fig. 3). Mutation of either of these conserved cysteine residues markedly alters the function of rhodopsin, muscarinic, and β -adrenergic receptors (Dixon *et al.*, 1987a; Karnik *et al.*, 1988; Fraser, 1989; Hulme *et al.*, 1990). The most highly conserved intracellular sequence is the aspartate-arginine-tyrosine triplet adjacent to TM 3 which has been implicated in signal transduction (see below). The arginine of this triplet is invariant, and the aspartate and tyrosine are conservatively replaced in several GPRs.

The amino termini of these proteins vary greatly in length. They range from as few as seven residues in the adenosine A_1 receptor to over 300 residues for the glycoprotein hormone receptors. Overall, there is little sequence homology among the receptors in the first extracellular domain. The amino termini of nearly all the GPRs contain consensus sequences (N-X-S/T) for *N*-glycosylation (Kornfeld and Kornfeld, 1985). Rhodopsin, the α_2 -adrenergic, the β_1 -adrenergic, and the β_2 -adrenergic receptors are all glycosylated at several of these sites (Hargrave, 1977; Strasser *et al.*, 1984; Benovic *et al.*, 1987b; Dohlman *et al.*, 1987; Regan, 1988). Glycosylation may contribute to the proper expression of membrane proteins (for review, see Kornfeld and Kornfeld, 1985). Deletion of the glycosylated domains of the β_2 -adrenergic receptor decreased the level of receptor expression but did not alter ligand binding (Dixon *et al.*, 1987b). Inhibition of glycosylation diminished muscarinic receptor expression (Liles and Nathanson, 1986). The thyroid-stimulating hormone receptor contains six potential glycosylation sites. Mutational analysis demonstrated that two of these sites are required for the expression of functional receptor (Russo *et al.*, 1991). Some receptors with short amino termini (the A_1 and A_2 adenosine and the α_{2B} -adrenergic receptors) do not contain amino-terminal asparagine glycosylation sites.

Phosphorylation and palmitoylation of carboxy-terminal sites can influence the signal transduction of some GPRs. Most GPRs contain potential phosphorylation sites in the third cytoplasmic loop and/or carboxyl terminus. For several receptors, phosphorylation by protein kinase A and specific receptor kinases mediates receptor desensitization (see Intracellular Coupling below, for discussion). Two adjacent cysteine residues in the carboxyl terminus of rhodopsin and one in the β_2 -adrenergic receptor are palmitoylated (Ovchinnikov *et al.*, 1988; O'Dowd *et al.*, 1989a). The hydrophobicity profile of the GPRs predict seven TM domains and thus three intracellular loops (Fig. 1). The palmitate on carboxy-terminal cysteine(s) would be expected to insert into the membrane, thereby forming an additional cytosolic loop which may influence receptor mobility (Findlay and Eliopoulos, 1990) or G-protein coupling (O'Dowd *et al.*, 1989a).

FIG. 2. Amino acid sequence alignment of the GPR superfamily. The putative transmembrane domains are enclosed in boxes. The precise boundaries of the TM domains are not known with certainty. Dashes have been introduced for the purpose of alignment. Amino acids omitted from nonconserved regions are indicated by numbers in parentheses.

1. Dictyostellium cAMP receptor (Klein et al., 1988)
2. Dog adenosine A2 receptor (RDC8) (Libert et al., 1989b)
3. Dog adenosine A1 receptor (RDC7) (Libert et al., 1989b)
4. Human m1 muscarinic acetylcholine receptor (Peralta et al., 1987)
5. Human m2 muscarinic acetylcholine receptor (Peralta et al., 1987)
6. Human m3 muscarinic acetylcholine receptor (Peralta et al., 1987)
7. Human m4 muscarinic acetylcholine receptor (Peralta et al., 1987)
8. Human m5 muscarinic acetylcholine receptor (Bennet et al., 1988)
9. Human beta 1 adrenergic receptor (Friedle et al., 1987)
10. Human beta 2 adrenergic receptor (Kobilka et al., 1987a)
11. Human beta 3 adrenergic receptor (Emorine et al., 1989)
12. Cow alpha 1 adrenergic receptor (Schwinn et al., 1990)
13. Rat alpha 1B adrenergic receptor (Voigt, et al., 1990)
14. Human alpha 2 C4 adrenergic receptor (Regan et al., 1988)
15. Human alpha 2 C2 adrenergic receptor (Lomasney et al., 1990)
16. Human alpha 2 C10 adrenergic receptor (Kobilka et al., 1987c)
17. Rat alpha 2 adrenergic receptor R20 (Lanier et al., 1991)
18. Drosophila octopamine receptor (Arakawa et al., 1990)
19. Human dopamine D1 receptor (Dearry et al., 1990)
20. Human dopamine D5 receptor (Sunahara et al., 1991)
21. Human dopamine D2 receptor (Grandy et al., 1989)
22. Human dopamine D3 receptor (Gilos et al., 1990)
23. Human dopamine D4 receptor (Van Tol et al., 1991)
24. Human serotonin 1d receptor (RDC4) (Hamblin and Metcalf, 1991)
25. Human serotonin 1a receptor (Kobilka et al., 1987b)
26. Rat serotonin 1c receptor (Julius et al., 1988)
27. Rat serotonin 2 receptor (Julius et al., 1990)
28. Human histamine H2 receptor (Gantz et al., 1991)
29. Human N-formyl peptide receptor (Boulay et al., 1990)
30. Human C5a anaphylatoxin receptor (Gerard and Gerard, 1991)
31. Human thrombin receptor (Vu et al., 1991)
32. Human thromboxane A2 receptor (Hirata et al., 1991)
33. Human IL-8 receptor (Murphy and Tiffany, 1991)
34. Guinea-pig platelet-activating factor receptor (Honda et al., 1991)
35. Cow endothelin 1 receptor (Arai et al., 1990)
36. Rat non-isopeptide selective endothelin receptor (Sakurai et al., 1990)
37. Mouse bombesin/gastrin releasing peptide receptor (Spindel et al., 1991)
38. Rat neuromedin B preferring bombesin receptor (Wada et al., 1991)
39. Human vasoactive intestinal peptide (Sreedharan et al., 1991)
40. Rat neurotensin receptor (Tanaka et al., 1990)
41. Rat bradykinin receptor (McEachern et al., 1991)
42. Mouse thyrotropin-releasing hormone receptor (Straub et al., 1990)
43. Human neurokinin A (SK) receptor (Gerard et al., 1990)
44. Rat substance P receptor (Yokota et al., 1989)
45. Rat neuromedin K receptor (Shigemoto et al., 1990)
46. Bovine adrenal angiotensin II type-1 receptor (Sasaki et al., 1991)
47. Human mas oncogene (angiotensin) receptor (Young et al., 1986)
48. Human lutropin-choriogonadotropin receptor (Frasier et al., 1990)
49. Human thyrotropin receptor (Libert et al., 1989a)
50. Human follicle stimulating hormone receptor (Mingos et al., 1991)
51. Human rhodopsin (Nathans and Hogness, 1984)
52. Human green opsin (Nathans et al., 1986)
53. Human red opsin (Nathans et al., 1986)
54. Human blue opsin (Nathans et al., 1986)
55. Odorant receptor F3 (Buck and Axel, 1991)
56. Odorant receptor F5 (Buck and Axel, 1991)
57. Odorant receptor F6 (Buck and Axel, 1991)
58. Odorant receptor F12 (Buck and Axel, 1991)
59. Odorant receptor I3 (Buck and Axel, 1991)
60. Odorant receptor I7 (Buck and Axel, 1991)
61. Odorant receptor I8 (Buck and Axel, 1991)
62. Odorant receptor I9 (Buck and Axel, 1991)
63. Odorant receptor I14 (Buck and Axel, 1991)
64. Odorant receptor I15 (Buck and Axel, 1991)
65. Human cannabinoid receptor (Matsuda et al., 1990)
66. Mouse Glucocorticoid-induced receptor (Harrigan et al., 1991)
67. Rat FC5R (Eva et al., 1990)
68. Human endothelial cell GPR (Hla and Maciag, 1990)
69. Rat testis G-protein coupled receptor 1 (Meyerhof et al., 1991a)
70. Rat RGHJP (Meyerhof, DNA and Cell Biology, in press, 1991b)
71. Human thoracic aorta GPR (Ross et al., 1990)
72. Cytomegalovirus (Human) GPR, US33 (Chee et al., 1990)
73. Cytomegalovirus (Human) GPR, US27 (Chee et al., 1990)
74. Cytomegalovirus (Human) GPR, US28 (Chee et al., 1990)

1 MGLDGNPANET
 2 MSTMGSW
 3 MPPAISAFQA
 4 MNTSAPPAVSPNITVLPAGKGPWO
 5 MNNSTNSSNLSALTSPYKTFE
 6 MTLHNNSTTSSPLFNNISSSWIHSPSDAGLPPGTVTHFGSYNVSRAGNFSNDGTTDDPLGGHTVWQ
 7 MANFTPVNGSSGNQSVRLVTSSSHNRYETVE
 8 MEGDSYHNATTVNGTPVNHQPLERHRLWE
 9 MGAGVLVLGASEPGNLSAAPLPDGAATAARLLVPASPPASLLPPASESPEPLSQOW
 10 MAPWPHENSSSLAPWPDLPPLAPNTANTSGLPVGPWE
 11 MGQPGNGSAFLLAPNRSHPADHDVTQORDEVW
 12 MVFLSGNASDSSNCTHPPPPVNI SK
 13 MNPDLDTGHNTSAPAHWELKDDNFTGPNQTSNSSTLPQLDVTR
 14 MASPALAAALAVAAAAGPNASGAGERGSGGVANASGASWGPFRGOYSAGA
 15 MDHQDPYSVOA
 16 MGSLOPDAGNASWNGTEAPGGGARATPYSLQV
 17 MGSLOPDAGNSSWNGTEAPGGGARATPYSLQV
 18 MPSADQILFVNVTTTVAALTAATAVSTTKSGNAARGYTDSDDDAGMTEAVANISGSLVEGLTIVTAALS- (35)
 19 MRTLNTSAMDTGLVVERDFSV
 20 MLPPGSGNGTAYPGQFALYQQLAQGNVAGSGAGAPPLGPS
 21 MCPILNLSWYDDLERQWRSRPNCSGDKADRP
 22 MASLSQLSSHLSNSTCGAENSTGASQARPH
 23 MGNRSTADADGILLAGRPAAGASAGASAGLAGQ
 24 MSPINOSAELPQEAENRSLNATETSEAWNPTLQAL
 25 MDVLSPOGQNNITSPAPFETGQNTGISDVTVSQ
 26 MVNLGNAVSLMHIGLLVWCFDII SPVAGIVDTTNSDGGRLFPQDGV
 27 MEILCEDNISLSSIPNSIMQLGDGPRLYHNDNFNSRDANTSEASNWTIDAENRNLSCGGLPPTCLSLHLQE
 28 MAPNGTASSFCIDSTACK
 29 METNSSLPINI SGGTPAVSAGYLF LD
 30 MNSFNITTPDYGHYDDKDTLDINTPVKTSNTLRVP
 31 MGPRRLLVAAACFSLCGPLLSARTRARRPESKATNATLDRSFLLLRNPNDKYEPFWEDEEKNESGLTEYRLVSINKSSPLQKQPAFISEDASGYLTSSWL
 32 MWPNGSSLCPCFRPTNITLERR
 33 MESDSFEDFWKGEDLSNYSYSTLPFLDLAAPCEPESLEIN
 34 MELNSSSRVDSEFRYT
 35 METFWLRLSFWVALVGGVSDNPESYSTNLSIHVDSVATFHGTELSFVVTTHQPTNLALP SNGSMHNYCPQOKTITSAPK
 36 MQSSASRCGRALVALLACGLLGWGEKRGFPAPQATPSLLGTKEVMTPTTKTSWTRGSNSLSMFRTAETVKGRVAGVPPRSFPQKRIEINKTFK
 37 MMAPNNCSHLALDVPFLSCNDTFNQSLSPPKMDNWFHPGF
 38 MPPRSLPNLSLPTASESELEPEVWENDFLPDSGTTAEIVIR
 39 MDLHLFDYAEPCNTSDI SWPCNSSDCIVVDVVMKPNMKNKSVLL
 40 MHLNSSVPQGTGEPDAQPFSCQSEMEATFLALSNGSGNTSESDTAGPNSDLVDNTDIYS
 41 MFNITTOALGSAHNGTSFEVNCPTTEWWSWLN
 42 MENDIVSEMNOTELOQAAVALEYQVVT
 43 MGTCDIVTEANISSGPESNITGITAFSMPSWQ
 44 MDNVLPMDSDLPFNI STNTSESNQFVQPTWQ
 45 MASVPRGENWTDGIVEVGTHTGNLSALGVTEWIALAQAGNFSALGLPATTOAPSQVRANLTNOFVQPSWR
 46 MI LNSSTEDGI KRI QDQCPKAGRHHYIFI
 47 MDGSNVTSFVVEEPTNISTGRNASVGNARHQP
 48 MKORFSPLOLLKLLLLLOAPLPRALRRLCPEPCN- (248) -LPTKEINFSHSISENFSKQCESTVRKSELGWDYEGFCLPKTPRCAPEPDAFNPCEDIMG
 49 MRPADLLQLVLLLDLPDLGGMGCSPPCECHQE- (318) -YVFEQEDEI IGFGQELKNPQEEETLQAFDSDHYDTTCGDSEDMVCTPKSDEFNPCEDIMG
 50 MALLVSLAFLSLGSGCHHRI CHCSNRVFLCQE- (266) -VDYMTQARGQRSSLAEDNESSY SRGFDMTYTEFDYDLNNEVVDVTCSPKPFDAFNPCEDIMG
 51 MNGTEGPNFYVPPSNATGVVRSFPEYPOYYLAEPWQF
 52 MAQQNSLQRLAGRHPQDSYEDSTOSSIFTYINSNSTRGPFEGPNYHIAPRWVYHLTSVW
 53 MAQQNSLQRLAGRHPQDSYEDSTOSSIFTYINSNSTRGPFEGPNYHIAPRWVYHLTSVW
 54 MRKMSEEFYLFKNISSVGPWDPQYHAIPVWAFYL
 55 MDSSNRTRVSEFLLGLFVENKDLQF
 56 MSTINQSSVTEFLLGLSROPQOQO
 57 MAWSTGQNLSTPGPFILLGFPGRSMRI
 58 MESGNSSTRFSSFFLLGFTENPQLHF
 59 MNNQTFITQFLLGLLPIPEEHQ
 60 MERRNHSGRVSEFVLLGLPAPAPLRV
 61 MNNKTVITHTFLLGLLPIPEEHQ
 62 MTRRNQTAISQFLLGLLPIPEEHQ
 63 MTGNQTLILEFLLGLLPIPEEHQ
 64 MTEENQTVISQFLLGLLPIPEEHQ
 65 MKSILDGLADTTFTITDILLYVGSNDIQYEDIK- (21) -SPFQEKMTAGDNSPLVPAGDTTNI TEFYKNSLSFKENEENI OCGENFMDMECFMI LNPSQO
 66 KVPPVLLFLSSVRATEQPVVTEHPSWEAALTGPNASSHFWANYTFSDQNFVGRRRYGAESQNPV
 67 MNSTLSFRVENYSVHYNVSENSPFLAFENDCHLPLAV
 68 MGPTSVELVKARRSSVSDYVNYDI IVRHYNYTGKLNISADKENSIX
 69 MKANNITTSALWLO
 70 MFPNGTAPSPTSPPSSSPGCGGVC SRGPGSGAADMEEGRNSSONGTLECGQS
 71 MAGNCSWEAHSTNCA: MCPQSEALELYSRGLTIEQIATLPPPA
 72 MTGFLFAIR
 73 MITSTNNQTLTQVSNMNHNTLNSTIEYQLFEYTR
 74 MPTTTTAEITTEFDYDEDATPCVFTDVLNQSK

1

GFWRLKLLRNHVTK
 —WLNSNLQNVIN
 —KVNQALRDATF
 —KVNTCLKIVNN
 —KVNRLHQTIVNN
 —KVNKQLKIVNN
 —KVNRLQTIVNN
 —KVNSOLKIVNN

—KTPR-LQTLN
 —KFER-LQTLVN
 —WTPR-LQTMN
 —CHRHLSVTH
 —CNRHRLKPTPN
 —TSRSLRAPON
 —TSRSLKAPON
 —TSRSLKAPON
 —TYKPLRIVON
 —RFYHLSKVIN
 —RSRHLRNMTN
 —REKALQTTIN
 —REKALQTTIN
 —TERALQTPTN
 —LTKRLHTPAN
 —LRSLOQVNA
 —MEKKLQNTN
 —LEKKLQNTN
 —LNRLRLNTN

-----FRMTHVTIT
 -----LEAKRTITPA
 -----LKMVKVKPAV
 -----QNKCMRNQPN
 -----YSRGVRSVID
 -----RLYP SKKNEIK
 -----KNKCMRNQPN
 GARQSSSHSTRSSFL
 -----CTVKSMMRNVPN
 -----LTNSTMSRVPN
 -----QAKTGYDTH
 RKKSLQSLQSTVAH
 -----LHKTNCTVAE
 -----RTKHMRTPTN
 -----AHRMRMTVTN
 -----AHKRMRTVTN
 -----AHRMRMTVTN
 -----FYMKLKTYS
 -----FRMRNPF

—TSRYKLTVP
 —TSHYKLVPR
 —TSQYKLTVP
 —QHKKLRTPLN
 —KFKKLRHPLN
 —KFKKLRHPLN
 —AYKKLRPLN

SDPC LHTPM
GDSRLHTPM
AHRC LQPMT
TQSH LHTPM
LDSQLHTPM
NHPT LHKPM
LDShLHTPM
LDShLHTPM
LDShLHTPM
LDShLHTPM

—HSRSLCRPSY
 —KNQRMHSATS
 —KQKEMRNVTN
 —KTKKFHRPMY
 —KLNRTLRTTF
 —RYAKMKTATN
 —FSIKRTPFSIY
 LTNRVLG—STPT—
 YYRRKKKSPSD—
 —WRRRIQCSGD—

2

NTAVNGGFCPLYA
ISTGFCACHN CL
INIGPRTYFHT CL

YLIMGH-WALGTIA CD
YTVIGY-WPLGPVW CD
YIIMNR-WALGNLA CD
YIIKGY-WPLGAVV CD
YIIMGR-WALGSLA CD

LUVWGR-WEYGSFF CB
 HILMKM-WTFGNFW CB
 LALTHG-WPLGATG CB
 FEILGY-WAFGRVF CN
 LEVLGY-WVLGRIF CN
 NELMAY-WYFGQVW CG
 NELIGY-WYFRRTW CB
 NEVNGY-WYFGKTN CB
 NEVMGY-WYFGKVV CB
 YSILGR-WEPGIHL CK
 AEIAGF-WPFGSFW CN
 AEVAGY-WPFGAF CB
 LEVUGE-WKF SRIH CB
 LEVTGGVWNS SRIH CB
 SEVQGAAWLLSPRI CB
 YTITHT-WNFQOIL CB
 YQVLNK-WTLGQVT CB
 AILYDYVWFLPRYL CB
 TILYGRYWLPSKL CB
 TYLSCK-WSFGKVF CN

KANGCHWPFGWFL C
 IVQHHWPFGGAA C
 YFSGSDWFGSEL C
 KLLAGRWPFQDNDFGVFL C
 KVNQWIFGTFL C
 YSNQGNWFLPKFL C
 KLLAGDWPF GAEMC
 QHAALFEWHAVDPCRL C
 KYLADRWLFGRL C
 RYFFDEWVFGKLG C
 LVQHNQWPMGELT C
 NFIWVHPWAFGDAG C
 IANNDFWLFGVEL C
 DSIYGS-WVYGYVG C
 YASHNIWYFGRAF C
 YAVHNWYGLFY C
 YGLHSEWYFGANY C
 TAMEYRWPGNYL C
 YALDYELSSGHYYTIV

DSQTKGQYYNHAIDWQTGSG-----C
DLYTHSEYYNHAIDWQTGPG-----C
DIHTKSQYHNYAIDWQTGAG-----C

-----TSLHGYFVFGPTG-----C
-----NQVYGYFVLGHPM-----C
-----NQVSGYFVLGHPM-----C
-----ASGNGYFVFGRHV-----C

KML VNIQTQNNVITYAG C
KVL ANHILGSGQAISFG C
KTL ATPARGGVSISLAG C
KML VNIYTQSKSITYED C
KLL QNMRSDQTSIPYGG C
KLMAGFIGSKENHGQLISFEA C
KLL QNIQSQVPSISYAG C
KLL QNMQSQVPSIPYAG C
KLL QNMQSQVPSISYTC C
KLL QNMQSQVPSIPFAG C

FHVFRKRDSPNV — F
 RFVNSTWYFGKM — C
 YTLNDHWVFGETM — C
 SGATTYKLTPAQWF —
 SAWRSRCTSMA — C
 TLLRHWFPGALL — C
 NMGTFLGSPDPYVR —
 NQWLLPAGVAS — C
 AKHHPKLSREVV — C
 LLDHNSLAS-VP —

1 IVITYGSFACWLWTICLAISY
2 FFACFVLVLTOSSIFSLIAIAI
3 MVACFVLILTOSILALLIAIV

4 LWLALDYVSNASVINLLLI SF
5 LWLALDYVSNASVMNLLII SF
6 LWLALDYVSNASVINLLVI SF
7 LWLALDYVSNASVMNLLII SF
8 LWLALDYVSNASVINLLVI SF

9 LHTSV DVL CVTASI ETLCVIAL
10 FWTSIDVLCVTASI ETLCVIAV
11 LNTSV DVL CVTASI ETLCALAV
12 VWAAVDVLCTCTASIMGLCII SI
13 IWAADVLCCTASI ILSLCAI SI
14 IYALVDVLCFTSSIVHLCAI SL
15 IYALVDVLCFTSSIVHLCAI SL
16 IYALVDVLCFTSSIVHLCAI SL
17 IYALVDVLCFTSSIVHLCAI SL
18 LNLTCVLCCTSSI INLCAI SL
19 IWAFTDMCTSTSI INLCVI SV
20 VWAFTDMCTSTSI INLCVI SV
21 IFVTLVMMCTSTSI INLCAI SI
22 VFVTLVMMCTSTSI INLCAI SI
23 ALMAVMMLCTSTSI INLCAI SV
24 IWLSSDITCTCTASI LHL CVIAL
25 LFLALDVLCTCTSSI IHLCAIAL
26 WVISLDFVSTASIMHLCAL SL
27 IWIYLDVLFSTASIMHLCAL SL
28 IYTSIDVMLCTSTSI INL FMI SL

29 FLFTVDDINLFGSVFLIALIAL
30 LPSLILLLMYASILLMTTASI
31 FVTAAFYCNMYASILLMTTASI
32 FMGVMMFFGLSPLLIGAMAS
33 VWSLKEVVFYSGILLLACTSV
34 LAGCLFFINTYCSVAFGLVITY
35 LFFFLQKS SVGVTINLNCALSV
36 LVFFIQKASVGVITLSCALSI
37 LIPFIQILTSVGVSVFTLTASA
38 LPIAQILTSVGVSVFTLTASA
39 VTHLIF SINLFGIFFLTQMSV
40 GYFFLRDCTATALNVAISLVS
41 VWNTHI YNLYSSICFLMWISI
42 CITYLQYLGINASSCSI TAFTI
43 FQNLFP ITAMFVYSYMTATAA
44 FHNFFP IAAALFASYSMTAVF
45 FQNFPP ITAVFASYSYSA-ALAV
46 IASASVSPNLYASVFLTCLSI
47 TLSVTF LFGYNTGLYLLTASV

48 TAGFFTVLASELSVYTLTVITL
49 TAGFFTVFASELSVYTLTVITL
50 AAGFFTVFASELSVYTLTAITL

51 LEGFFATLGC EIALWSLVLAI
52 LEGYTVSLCGITGLWSLAIISW
53 LEGYTVSLCGITGLWSLAIISW
54 LEGFLGTVAGLVTGWSLAF

56 TQIYFLFVAVGNDNDFLLAVMSY
57 TQMYFVFSLGCTEYFLLAVMAY
58 SQMCVFLVFAELGNFLLAVMAY
59 AQTYFFMFVGDMSFLLVAMAY
60 TQIYFLFGLGCTEVCVLLAVMAY
61 TQIFFFFLLFGYLGNFLLVAMAY
62 AQTYFFLFFGDLGNFLLVAMAY
63 TQIYFFMFVGDMSFLLVAMAY
64 TQLYFYLYFADLESFLLVAMAY

65 FKLGGVTASFTASVGSFLITAI
66 VSRFAQCVSLHVSALLITAIIV
67 LNPVQVCVSTIVISFLVLIIV
68 LREGSMFVSLSVFLSIIAII
69 FMSCVLLVFTHIMSLLAIIV
70 LVLSDAVNMFTSIYCLTVLSV
71 VSRVIGLCTFFAGVSLTPAIS
72 FLSVIYYSCCTVGFATVALIAA
73 GLNACVYISCLFAGCVFLINLS
74 LLTACFPVAMFASLCFIEITAI

—MLIVKREPEPELFEX—
 DRYIAIRIPLRYNGLVGTGR—
 DRYLRVKIPLRYKTVVTPRR—

DRYFSVTRPLSYRAKRT-PRR--
DRYFCVTKPLTPVKRT-TKM--
DRYFSITRPLTYRAKRT-TKR--
DRYFCVTKPLTPARRT-TKM--
DRYFSITRPLTYRAKRT-PKR--

DRYLAI TSPF RYQSL LTRAR
 DRYFAI TSPF KYQSL LTKNK
 DRYLAVNIP LRYGALV LTKRC
 DRYIGVSY PLRYPTI VTQKR
 DRYIGVSY QLQYPT LVTQRK
 DRYWVSQVAEY NLYKRT PPR
 DRYWVSRALEY NSKRT PPR
 DRYWSI TQAI EYNL KRT PPR
 DRYWSI TQAI EYNL KRT PPR
 DRYWAI TDPINAY KRTVGR
 DRYWAI SSFPRYER KMT- KA
 DRYWAI SRPF RYKRM TQM
 DRYTAVAMP LYN- TRYSSKR
 DRYTAVMPVY OHGTQ GSSCR
 DRYVAVAVP LRYN- RQCSSR
 DRYWAI TDAL EYSKRTAGH
 DRYWAI TDP DVN KRT PPR
 DRYVALRNP IEHSF- SRTK
 DRYVAIQNP IHHSF RNSRTK
 DRYCAVMDP LRYPLV LTPVR

DRVCVCLHPVWTONHRTVSLAK--
DRFLLVFKVPMQCNFRAGLR--
DRFLVVPVPMQSLWSRLTRGL--
ERYLGI TRFSPRAVASQRR--
DRYLAIIVHATRLITOKRHLVK--
NRFQAVKYP IKTQAQATTKR--
DRYRAVASWSRVOGIGIPLV--
DRYRAVASWSRIGIKGVPK--
DRYKAIVRPMDIQASHALMK--
DRYRAIVNPMQMTSGVVL--
DRYLSITYFTNPSRSRKMVRR--
ERYLAICHFFKAKTLMSSRSTK--
DRYLALVKMTSGRMGRVR--
ERYIAICHPIKAQFLCTFSR--
DRYMAIVHPFQRLSPASTK--
DRYMAI IHPLOPLRSATK--
DRYMAI IDPLKPLRSATATK--
DRYMAIIVHPMKRSLRMTL--
ERCRLSVLYPIWYRCHRPKY--

ERWHTITYAIHLDQKRLRLH---
ERWYAITFAMRLDRKIRLRLH---
ERWHTITHIMOLDCKVOLRLH---

RYVVVCKPMSNFRFGEN-----
RWVVVCKPFGNVRFDAK-----
RWLVVCKPFGNVRFDAK-----
RYIVICKPFGNFRFSSK-----

RYVAICHPMHYTVIMNYK---
RYVAICHPLHYTTIMTRO---
RYLAICLPLRYGTMTPG---
RYVAXCHPLCTYVIVNHR---
RYVAICFLPHYTSIMSPK---
RYVAICHPLHYPVIVSSR---
RYVAICFPLHYTNIMSHK---
RYVAICFPLHYMSIMSPK---
RYVAICFPLRYTTIMSTK---
RYVAICFPLHYMSIMSPK---

```

RYISIHRLPLAYKRIVTRPK---
RHQVIMHPLKPRISITKG---
RHQLIINPRGWRPNNRH---
RYITLKMMLKLGSSNNFR---
RYLRVKLTVRYRTVTTQRR---
RYVAVVHPIKAARYRRPT---
RCVSVIFPMWYWRRRPKR---
RYRVLHKRTYARQSYR---
RYCVI VWGVELNRVNRNKR---
RYATVYMYR---RPVK

```

—YYLLCWGLPLISTIVMLA—
—AKGIIAVCWLSFAIGLTPML—GW
—AAVA IAGCWT I SEWGLTPIE—GW

-AALMIGLAWLVSFVL-WAPAILFW
-AGMIAAAWVLSFIL-WAPAILFW
-AGVMIGLAWVISFVL-WAPAILFW
-AGLMIAAWVLSFVL-WAPAILFW
-AGIMIGLAWLISFIL-WAPAILCW

-ARGLVCTVWAIISALVSFLPIIMHW
 -ARVITIMMWIVSGTSFLFIPQIMHW
 -ARTAVLWVWVSAVSFAPIMSQW
 -GLPALLCWALSIVISIGPLF-GW
 -AIIALLSVWLTSTVISIGPLI-GW
 -VKATIVWVLISAVISFPPVLSLY
 -IKCIITLWVLLAAVISLPPLIYKG
 -IKAITITVWISAVISFPPLIISI
 -IKAITVWVWISAVISFPPLIISI
 -VLLISGWLLSLLISPPLI-GW
 -AFILISVANTLSVLISFIPVQLSW
 -ALVMVGLATLSLISFIPVQLNW
 -VTVMISVWVLSFTISC-PLLFGL
 -VALMISWVWIAFASVC-PLLFGL
 -QLLLIGATWLLSAWAA-PVLCGL
 -AATMIAWVAISICISIPPLF-LW
 -PRALISLWLGFLSIPPM-GW
 -AIMKIAWVAISIGVSV-PIPVIG
 -AFLKIAVWITISVGSIM-PIPVIG
 -VAISVLWISITISLFSLISLWLG

--KVIIGPWMAILL--TLPVII--
 --AWTACAVAGWIALLL--TIPPSFL--
 --ASTFLTAWALIAAG--V--PFLV--
 --AWATVGLVMAAIALGLLPILL--GV--
 --FICLSINGWLSILL--ALPVL--
 --GTALSIVTWALIVAA--ASFYFLVM--
 --TAIEIVSWILSFILL--ALPEAIGF--
 WTAVEIVLWVSVVL--AVPEAIGF--
 --ICLKAAVIVSVML--ALPEAVE--
 --WTSVAGVWVSVVL--AVEPEAVE--
 --AVCILWVLAFVCV--SLDPTYYL--
 --KFISAIWLASALL--AIPMLFT--
 WAKLYSLVSWCTILL--SSPMFLF--
 --AKKIIIFWAFSTSYIOMFLVLD--
 --AVIAGVILVALAL--AFPOCFY--
 --WVIFVIWVLALL--ASPOGYY--
 --IVGSIWTLAPLL--AFPOCYL--
 VAKVTCIIWILAGLA--SLPTIHR--
 OALVLCALIAWISLCTVTE--VUM--

-A I L I M L G G W L F S S L I A M L P L V G V S
-A C A I M V G G W V C C F L L A L L P L V G I S
-A A S V M M C W L E A F A A A L E R I E G I C

HAIMGVAFTHVMALA-CAAPPLAGW
LAIVGIAFSWIWAAV-WTAPPIFGW
LAIVGIAFSWIWSAV-WTAPPIFGW
HALTVVLATWTIGIG-VSIPFFGW

LCGLFLVLSWIVSVLHALFQSIMML
LCMLLVGGSWVWVNMNCLHTLLMA
LAVRLAGSLWLCGFSAITVPTLIA
LCI LLLLLSWVISIFHAFIQSLVIL
LCTCLVLLMLITTSHAMHTLLAA
LCVQMAAGSNAGGFGISMVVKFLSI
LCTQLLVFWIMTSSHAMHTLLAA
LCVSLVVLWSVLTTFFHAMHTLLMA
FCASIVLLMLMTTHALLHTLLAI
LCSTVLLGSLTSTTHALLHTLLAI

-AVVAFCLMWTIAIVIAVLP LL-GW
 -VYIIAVINWMTTF-SIPHAIC-
 -AYIGITVIVWLAVAS-SLPFVYI-
 -LFLLISACWISLIILGGLPIM-GW
 -IWLFLGLCLWLSFVIGLTPMF-GW
 -VAGVNVIGVWLISLV-LPIVWFS
 LACKWAGLWLLSFL-TSISHNYF-
 STYMI LLLTWLAGLI-FSVPAVYT
 DCLWVIE-WILAVL-MGNPHYIMY
 DCCSFTTETETETETETETETETET

-----KNTVQFVGN
-----NNCSQPKGRNY SQ
-----NRLCEAOBAWANGSGCERUE

QYLVGERTMLAG
QFIVGVRTVEDG
QYFVGKRTVPPG
QYFVGKRTVDPN
QYIVGKRTVPIA

-----WRAESDEAR
 -----YRATHQEI
 -----WRVGADAEAQ
 -----RQPAPEDQT
 -----KEPAPNDK
 -----RQPDGAAYP
 -----DQGPQPRGR
 -----EKKGGGGGQPAEP
 -----EKKGAGGGQPAEP
 -----NDWPDEFTSAT
 -----HKAKPTSPSDGNATSLAETID
 -----NRDQAASWGGGLDLPNN (20)
 -----NNADQN
 -----NTGDPT
 -----NDVRGRDPA
 -----RQAKAQEEMS
 -----RTPEDRSDDP
 -----LRDSKVFVNNT
 -----LQDSSKVFKEG
 -----NSRNETSKGNHTTS

RVITVPGKTGTV
RVVREEYFPFKV
KEQTIQVPLNIT
GRYTVQYPGS
FRRTVYSSNVSP
DSTMVSNKAGSNT
WMVPFEYKGAQHR
DITSDYKGRPLR
SDLHPFHVKTNTQFI
SEVARI-GSSDSSST
KVTTSASNNET
MGLQNRSGDGTHPGGL
TMDYREEGHNV
LNISTYKNVAVV
STVTMDQCAT
STTEIMPSRV
SKI KVMPPGR
NFFIENTNIT
CYDEFEHJEN

NYMKVS
SYAKVS

_____SRYIPEGLQC
 _____SRYWPHGLKT
 _____SRYWPHGLKT
 _____SRFIPEGLQC

ALPFCTHLEIPHY
 RRSFCADNMIPHF
 RLSFCGSRVINHF
 QLTFCGDVKIPHF
 RLSFCENNVVLNF
 RLSYCGPNTINHF
 RLSFCENNVLLNF
 RLSFCEDSVIPHY
 RLSFCEKNVILHF

_____NCKKLQS
 _____QKLTFTKYSEIDVRS
 _____OILTDEPFQNVLSAIFDKY
 _____NCISALS
 _____NRKVLTLELQNSSTL
 _____RTAANSOGTV
 _____CMFLGHEASGT
 _____TVVMHHDANDNTNNTNGHA
 _____SHTNN

1	WCWIGVSFTGCRFG	-LFYPFLFIWASAVLVGLT-	SRITYVVIHNGVSDN
2	GCCEGQVACLFEDVPMN	YMYVYNF FAFVLVPLLLMLGVYL	RIFLAARQLKQMESQPLPGERARSTLO
3	KCEFEKVISME	YMYVYNF FVWVLPPLLLMLVLYL	EVFYLIRRLQCKKVSASSGDPQKYTG
4	QCYIQFLSQP	IITFTGTAFAFYMPVIMCTLYW	RIYRETNAREIAALQSGSETPGKGGSSSSSSERSQPGAEGSPETP
5	ECYIQFFSNP	AVITGTATAAFYLPVIMTVLYW	HISRAKSKRIKKKKPEVANQDPVPSLVQGRIVKPNMNMPSDD
6	ECYIQFLSEP	TITFTGTAIAAFYMPVIMTVLYW	RIYKETEKRITKELAGLOASGTEAETENFVHTGSSRSCSSYLEQOO
7	QCYIQFLSNP	AVITGTATAAFYLPVIMTVLYI	HISLASRSRVHKKRPEGPKKAKTALFLKSPIMKQSVKPPPGEA
8	ECYIQFLSEP	TITFTGTAIAAFYIPVIMTVLYC	RIYRETEKRTKDLADLQSDSVYKAERKRAHRALEFRSLCPCRPRT
9	RCYNDPKCCDFVNR	AYAIASSVVSFYVPLCIMAFLYL	RVFREAOQKQVKKIDSCERRFLGGPARPPSPSPVFPAPAPPGPFRP
10	NCYANETCCDFTNQ	AYA-ASSAVSFYVPLVIMTVLYS	RVFOEAKRQLQKIDKSEGRFHVQNLQVQDGTGHTGLRSSSKFCL
11	RCHSNPRCCAFASNM	PYVLLSSVSFYVPLVIMTVLYA	RVFVATROLRLRLGEILGRFPPEESPAPSRSLAPAPVGTGAPPEG
12	ICQINEEP	GYVLFSAIGSFYVPLTILVMYC	RVYVAKRESRGLKSLKTDKSDSEQVTLRIHRKNAQVGGSGVTSAR
13	ECVTEEP	FCALFCSLGSFYIPLAVILVMYC	RVYIVAKRTTKNLEAGVMKMSKELTLRIHWSKNFHEDTSLSTK
14	QCGINDET	WYLLSSCIGSFAPCLIMGLVYA	RIYRVAKRRTTLSEKRAVPGPDGASPTTENGILGAAGEARTGTAR
15	QCKINQEA	WYLLASSIGSFAPCLIMGLVYL	RIYLIAKRSNRRGPRAKGGPGOGESKQPRPDHGGALASAKLPALAS
16	RCEINDOK	WYVSSCIGSFAPCLIMGLVYV	RIYQIAKRRTRVPPSRRODPAVAAPPGGTERPNGLGPERSAGPGC
17	SKINDOK	WYVSSCIGSFAPCLIMGLVYV	RIYQIAKRRTRVPPSRRODPAVAAPPGGTERPNGLGPERSAGPGC
18	PCELTQRI	GYVYSSIGSFYIPLAIVTVYI	EIVFVATRRRLRERARANKLNTALKSTELFMANSSPVAASNSGSK
19	NCSSLSR	TYAISSVISFYIPLAIVTVYT	RIYRIAQKQIRRIALALERAHVAKNCOTTGNKPVCESSQFESSFKM
20	-CDSSLNR	TYAISSVISFYIPLAIVTVYT	RIYRIAQKQIRRIALALERAHVAKNCOTTGNKPVCESSQFESSFKM
21	ECIIANP	AFVYSSIVSFYVPLVIMTVLYI	KIYIVLRRRRRKRVNTRKSRAFAHRLAPLKGNCTHPEPMKCTVI
22	VCSISNP	DFVYSSVVSFYVPLVIMTVLYA	RIYVVLKQRRRKRIITRONSQCNSVRPGFPQOSTSLPDPALHLEKR
23	VCLERDR	DYVYSSVVSFYVPLVIMTVLYA	ATFRGLQWVARRAKLHGRAPRRSGPGPPSPPTPAPRILPDQPC
24	DCLVNTSQ	SYTIYSTCGAFYIPSVLLIILYG	RIYRAARNRIINPPSLYCKRFTTAHLITGASGSLCINSSLSHEGH
25	ACTISKDH	GYTIYSTIFAFYIPVLLIMVLYG	RIFRAARFRIKTVKKVEKTGADTRHGASPAPOPKKSVNGESGSRN
26	TCVINDPN	FVLIGSFVA-FFIPLTIVITFYF	LTIYVLRQTIMLLRGHTEELANMSLNLNCCCKKNGEENAPN
27	SCLLADDN	FVLIGSFVA-FFIPLTIVITFYF	LTIKSLQKEATLCVSDLSRAKLASFSLPQSSLSSEKLFQRSIHR
28	KCKVQVNE	VYGLVDGLVTFYLPVLLIMVITYY	RIFKVARDAKRNHISWKAATI
29	ACTFNFSPWINDPKER	INVAVAMLTVRGIIRFIIGFSAPM	SIVAVSYGLIATKINKQGI
30	LCGCDYSHDKRRER	AVAVIRLVGLFVPLVIMTVLYI	FILLRTWSRRA
31	TCHDVNLNETLLEGGYA	YFSAFSAVFFVPLIISTVCYVS	IIRCLSSSAVANRSKKS
32	WCFLLTGAESGDVAFG	LLFSLMGLSGLSVLNLNIVSA	TLHHVYHGQEAQQRPR
33	ACYEDMGNNYANRM	LLRIIPQSPGFIVPLIMVLYG	TLRLTFKAHM
34	RCFEHYEKGSKPV	LIHICIVLGFVPLVIMTVLYI	VIIHTLLRGPVKQORNA
35	TCMLNATSK-FMEFYQDV-KD	-WMLFGFYPCMLVCTAIFYTL	MTCEMLNRNGSLRIALS EHL
36	VCMLNPFQKTAFFQFYKTAARD	-WMLFAFYFCLPLAITAIFYTL	MTCEMLNRNGSLRIALS EHL
37	SCAPYPHSNELHPK	IHSMAFLVYVPLIISTVCYVS	IARNLIQSAYNLEPVEGNIHVKKOI
38	ACIPYPQDELHPK	IHSVLIIVYVPLIISTVCYVS	IATKLIRSAHNLPEGYNEHTKKOM
39	YCRFPYPEHSIKWLI	SMELVSVVGLFAVFPFIIAVFYF	LLARAIASSD
40	VCTPIVDTATVK	VVIQNTFMFLFPLVIMTVLYI	VIANKLITVMHQAAEQGRVCTVGTNGHLEHSTFNMTEPGRV
41	TCVIVYPSRSWEV	FTMLNLNVLGFLPLIISTVCYVS	IMQVLNRNMKKFKEVO
42	SCGYKISRNYS	PIYIMDFGVFYVPMIATVLYG	IARILFLNPIPSDPKENSIMWKNDSIHONKNLNLNA
43	KCVWAPEDSGGKTL	LYHLVIALIYFLPLVIMTVLYA	VIGLTLWRAVPGHQAAGANLRL
44	VCMIWEPHFNRYEK	AYHICVTVLIYFLPLVIMTVLYA	VVGITLWASEIPGSSDRYHEQV
45	LCYV-WPEGPKQHF	TYHIVIIIVYCFPLIIMTVLYI	IVGITLWAGEIPGDCDYHEQL
46	VCAFHYESQNSTLPV	GLGLTKNIGLFLPLIISTVCYVS	LIWKTLLKAYEIQKNKP
47	DCRAVI	IFIAILSFLVFT-PLMVSSSTIL	VVKIRKNTWAS
48	ICLPMOVETTLQ	-VYIITLILNVAFLIICACYI-	KIYFAVRNPELMATN
49	ICLPMOTETPLAL	-AYIVFVLTNLIVAFVIVCCCYV-	KIYITVRNPQYNPGD
50	ICLPMIDSPLSQ	-LYVMSLLVNLVAFVIVCGCYT-	HIYITVRNPQYNPGD
51	SCGIDYTLKPEVNE	SFVIYMFVHFTIPMIIFFCYG	QLVFTVKEAAQQQESATTO
52	SCGPDVFGSSYPGVO	SYNIVLMVTCCTPLSIIVLCYL	QVWLAIIRAVAKQKESESTO
53	SCGPDVFGSSYPGVO	SYNIVLMVTCCTPLSIIVLCYL	QVWLAIIRAVAKQKESESTO
54	SCGPDWYTVGTYKRS	SYTWFLIFCFIVPLSLICFSYT	QLLRALKAVAAQQQESATTO
55	FCEPNOVIQLTCSDAFLND	LVYIYFTLVLLATVPLAGIFYSYF	KIVSSIC
56	FCDGTPLLKLSCTDTHNE	LMILTEGAVVMVTFVCLISYI	HITCAVL
57	FCDISPWIVLSCDTQVVE	LVSFGIAFCVILGSGITLVSYA	YIITITII
58	FCEINQLSQTCSDNFPH	LIMNLVPMVLAISFSGILYSYF	KIVSSIH
59	FCDLFVLLKLACSDTYINE	LMIFIMSTLLIIIPFLIVMSYA	RIISSIL
60	FCDVSPLLNLSCDMSTAE	LTDVLAIFILGPLSVTCASYM	AITGAVM
61	FCDLFVLLKLACSDTYINE	LMIHIMGVIIIVIPVLIIVSYA	KIISIL
62	FCDMSTLLKVCSDTHDNE	LAIFILGPIVPLVPLIIVSYA	RIVSSIF
63	FCDISALLKLSCSDIYVNE	LMYIILGGLIIIPFLIVMSYV	RIFFSIL
64	FCDISPLKLSCSDTHVNE	LVIFVMGGLVIVIPVLIIVSYA	RIVASIL
65	VCCDIFPLIDGTLYM	FWIGVTSVLLFIVYVYMYILW	KAHSHAVRMIORGTKSIIHTSEDGKVQVTRPDQA
66	ICLDPPPEPADLFWK	YLDIATFILLYLLPLFISVAYA	RVAKKLWLCNTIGDVTTEQYLALR
67	VCFDKFPSDSHRL	SYTLLIVLQYFGPLCFIFICYF	KIYIRLKRNNMMKIRDSKYRSS
68	SCSTVLPYLYKH	-YILFCTTVFTLLLSIVILYC-	RIYSLVTRSRRLTFRKNISKASRS
69	SCHFRSVVGLD	-YMFPSFITWILPLVVMCIYLD	IFYIIRNKLSONLTGFRTRAFCY
70	ACNMMPPEPAQRWLV	GFVLYTFIMGFLLPVGACILCYV	LIIAKRMVWALKAGWQORRR
71	ACLNMD	ISLGIILFFLFC-PLMPLCLAL	LHVECRARRRO
72	TCVLYFAVEEVHTVLL	SWKVLITMVGAAAPVIMTVLYA	FFYSTVORTSQ
73	ECVGMFANETSQWFPV	FLNTKVNICGYLAPIALMAYTYN	RMVRFIINYVG
74	QCMIDYDYLEVSYPI	ILNVEIMLGAFVILSVISYCYI	RISRIVAVSQS

1	-----KEKHLTYQFK	LINYIIVFLVCHVFAVNVNIVNGL	NMFPPALNHLTYL
2	-----KEVHAAS	LAIIVGLFALCWLPLHINCITFF	-----CPECSHAPIM
3	-----KELKIAKS	LALILFLFALSWSLPLHINCITLF	-----CPSCKRPSI
4	(83) -KGQKPRGKEQLAKRKTFSLVKEKKAART	LSAIIILAFILTWTPYNNIMLVSTF	-----CKDCVPET
5	(110) -K-IVKMTK-OPAKKKP-PPSREKKVTRT	IIAIIILAFIITWAPYNNMVLINTF	-----CAPCIPNT
6	(166) -KRFALKTRSOITKRKMSLVKEKKAQOT	LSAIIILAFIITWTPYNNIMLVNTF	-----CDSCIPKT
7	(113) -K-FASIAKNQVRKKROM-AARERKVTRT	IFAIIILAFILTWTPYNNMVLNTF	-----COSCIPTD
8	(155) -KGLNPNPSHQMTKRKMSLVKERKAAQOT	LSAIIILAFIITWTPYNNIMLVSTF	-----CKDCVPVT
9	-AAAAATAPLANGRAGKRRPSRLVALREQKALKT	LGIIMGVFTLCWLPFFLANVVKAF	-----HRELVPDR
10	-----KEHKALKT	LGIIMGTFTLCWLPFFIVNIVHVI	-----QDNLRKE
11	-----VPACGRRPAPLIPIREHRLACT	LGLIMGTFTLCWLPFFLANVLRAL	-----GGPSLVPGP
12	-----KNKTHFSVRLKFSREKKAART	LGIIVGCFVLCWLPFFIVMPLGSF	-----FPDFRPSET
13	-----AKGHNPSSIAVKLFKFSREKKAART	LGIIVGCFVLCWLPFFIALPLGSL	-----FSTLKPDA
14	(77) -FLSRRRARSSVCRKVAQAREKRTFV	LAVVGVFVLCWLPFFFIISLYGI	-----CREACQVPGP
15	(106) -GRGVGAIGQWRRRAHVTREKRTFV	LAVVIGVFLCWFPPFFSYSLGAI	-----CPKHCRVPHG
16	(84) -GRGRSASGLPRRRAGAGGONREKRTFV	LAVVIGVFWCWFPPFFTYTLTAV	-----GCSVPR
17	(84) -GQGEERAGGAKASRWGRONREKRTFV	LAVVIGVFWCWFPPFFTYTLTAV	-----GCPVPYQ
18	(167) -KKTSGVNQFIEEKQKISLSKERRAART	LGIIMGVFVICWLPFFIMYVILPF	-----CQTCCTNK
19	-----SFKRETKVLKT	LSVIMGVFVCCWLPFFIINCILPF	-----CGSGETQPPCIDSN
20	-----KETKVLKT	LSVIMGVFVCCWLPFFIINCMPFF	-----CSGHPECPAGFPVSET
21	(91) -PNGKRTSLKMSRBKLSQKEKKATQM	LAIVLGVFIICWLPFFIITHILNIH	-----CDCNIPPV
22	(47) -SNGRLSTSLKLPQPRGVPLREKKAQOM	VAIVLGAFIVCWLPPFLTHVLAATH	-----COTCHVSP
23	(29) -ALPPQTPPTQRRRRRAKI TGRERKAMRV	LPVVGAFILCWTFFVWHITQAL	-----CPACSVPR
24	(10) -NHVKIKLADSALERKRI SAARERKATKI	LGIILGAFIICWLPFFVSLVLPPI	-----CRDSCNHPA
25	(57) -ASFERNKRNAAEKRMALAREKVTKT	LGIIMGTFTLCWLPFFIVALVLPF	-----CESSCHMTL
26	-NPNPDQKPRKKKEKPRGTMOA INNEKKAQKV	LGIIVFVFLIMCWFPIITNLSVL	-----CGKACNOLMEK
27	-----EPGSYAGRTMOSISNEQKACKV	LGIIVFVFLVMMCPFFITNIMAVI	-----CKESCENVIGA
28	-----REHKATVT	LAIVMGAFIICWFPIITAFVYRGL	-----RGDDAINEV
29	-----IKSSRLPRV	LSFVAAFFLCWSPYQVVALIATV	-----RIRELQGMKEYIGI
30	-----TRSTKTLKV	WAVVASFFIFWLPYQVTGIMMSF	-----LEPSSPTFLINK
31	-----TNRCFNSTV	ALFLSAAVFCIFIICFGPTNVLLI	-----AHYSFLSHSTTEAAYF
32	-----DSEVEMAQ	ILGIMVASVSWLPPLIVIAQTVL	-----RNPAMPAPQLSRTEKE
33	-----GQKHRRMRV	IFAVVLIFLLCWLPLNVLADTL	-----MRTQVIQETCERRNHIDR
34	-----EVRRLALWM	VCTVLAIVFVFCVPHHMVQLPWTI	-----AELGMWSSNHQAIND
35	-----KORREVAKT	VFCVLVIFALCWFPLHLSRILKKT	-----VYDEMOTNRCELLSFLIL
36	-----KORREVAKT	VFCVLVIFALCWFPLHLSRILKKT	-----LYDQSNPORCELLSFLIL
37	-----ESRKRLAKT	VLFVGLFALCWFPLNHHVLYRSY	-----HYSEVDTSMHIV
38	-----ETKRRLAKI	VLFVGCFFVCFWPNHHVLYRSY	-----NYKEIDPSLGMMI
39	-----QEKHSRKI	IFSYVWVFLVWLPYHVAVLIDIF	-----SLIHYIPFTRLEHALFT
40	-----QALRHGVLV	LRAVVIAFVWLPYHVRRLMFCY	-----ISDEQWTFIFDFYHY
41	-----TEKKATVL	VLAIVGLFVLCWFPOISTFLOTL	-----LRLGLVSGCMNERAVDI
42	-----SSRKQVTKM	LAVVILFALLAMPYRTLVVNSF	-----LSSPFOENMK
43	-----QAKKFFVKT	MVLVWVFAICWLPYHLYFILGSF	-----QEDIYCHKFIOQ
44	-----SAKRKVVKM	MVWVCTFAICWLPFHVFFLLPYI	-----NPDLYLKKFIOQ
45	-----KAKRKVVKM	MIWVWVFAICWLPYHLYFILTAL	-----YQQLNRWKYIOQ
46	-----RKDDIFKI	IIAIVLFFFFSWPHNIFTMDVL	-----IQGLIRDCIKEDIVDT
47	-----HSSKLYIV	IMVTIIIFLIFAMPRLLYLLYE	-----YWTFTGN
48	-----KDTKIAKRM	AILIFTDFT-CHAPISFFAISAAP	-----KVPLITVNSK
49	-----KDTKIAKRM	AVLIFTDFI-CHAPISFYALSAIL	-----MKPLITVNSK
50	-----SDTRIAKRM	AMLIFTDFL-CHAPISFFAISAAL	-----KVPLITVSKAK
51	-----KAEKEVTRM	VIIIMVIAFLICWVPYASVAFYIFT	-----HOGSNFGPI
52	-----KAEKEVTRM	VVMVLAFCFCWGPYAFACFAAA	-----NPGYFPHPL
53	-----KAEKEVTRM	VVMIFAYCVWGPYTFACFAAA	-----NPGYAFHPL
54	-----KAEREVSRLM	VVMVGSFCVCYVPYAFAMVMN	-----NRNHGIDL
55	-----AISVHGKYK	AFSTCASHLSVSLFYCTGLCVYL	-----SSAANNSSQASA
56	-----RVSSPRGGWK	SFSTCGSHLAVVCLFYGTIVAVYF	-----NPSSSHIAGROM
57	-----KIPSARGHR	AFSTCASHLTVLWYGSTIFLHV	-----RTSVESLDTK
58	-----SISTVQGYK	AFSTCASHLSIVSLFYSTGLCVYL	-----SSAVVQSSHAA
59	-----KVPSTQICK	VFSTCGSHLSVSLFYGTIIGLYL	-----CPAGNNTVKEM
60	-----RIPSAAGRHK	AFSTCASHLTVVIIFYAASIFTYA	-----RPKALSFTDNK
61	-----KVPSTQSHK	VFSTCGSHLSVSLFYGTIIGLYL	-----CPSGDNFLKGS
62	-----KVPSSQSHK	AFSTCGSHLSVSLFYGTIIGLYL	-----CPSANNSEVKET
63	-----KFPISQDIYK	VFSTCGSHLSVSLFYGTIIGLYL	-----CPSGNNSTVKEI
64	-----KVPVSRGIHK	IFSTCGSHLSVSLFYGTIIGLYL	-----CPSANNSTVKEI
65	-----RMDIRLAKT	LVLILVLLIICWGPLLAIMVYDVF	-----GKMKLIK
66	-----RKKKTIVKM	LVLVWVLFALCWFPLNICYVLLS-	-----SKAIHTNNA
67	-----ETKRINV-M	LLSTVAFVAVCWLPPLTIFNTVFDW	-----NHQIIATCNHNL
68	-----SENVALLKT	VIIIVLSVFIACWAPLFIILLIDVG	-----CKVKTCDILFR
69	-----REFKTAKS	LFLVLFALCWLPLSIINFSYF	-----NVKIPET
70	-----SERKITLM	VMMVWVFCWMPFFYVQLVNPF	-----AEQDDAT
71	-----RSAKLNHV	VLAIVSVFLV-SSIYLGIDWFLFW	-----VFQIPAPF
72	-----KORSRTLF	VSVLLISFVALQTPYVSLMIFNSY	-----ATTAWPMQCEHLTLRRT
73	-----KWHMOTLHV	LIVVVSFASFVFPFNIALFLESI	-----RLLAGVYNDTLQNVIF
74	-----RHKGRIVRV	LIAVVLVFIIFWLPYHLLTFVDTI	-----KLLKWISSSCFEFSLKR

7

1	SVSHGFASVTFIYNPNIM-WRYF	GAKILTVFTFFGYFTDVQKKLEKNKNNNPSPYSSSRGTSKTMGGHPTGDDVQCSSDMEQCSLERHPNMV- (63)
2	LMYLTIVLSHTNSVNPFI-YAYR	IREFRQTFRKII RSHVLRREPFKAGGTSARALAAHGSDEQISLRNLGHPGCVWANGSAPHERRRPNGT- (50)
3	LMYIAIFLTHGNSAMNPV-YAFR	IQKFRVTF LKINWHDHFRQCTPPVDEDPPEEAPHD
4	LMELGYWLCYVNSTINPM-C-YALC	NKAFRDTFRLLLCFWDKRRWRKIPKRPQSVHRTPSQC
5	VMTIGYWLCYINSTINPAC-YALC	NATFKKTFKHLIMCHYKNIATR
6	YNNLGYWLCYINSTVNPVC-YALC	NKTFRTTFKTLLLCQCDKRRKQOYQORQSVIFHKRVPEQAL
7	VNSIGYWLCYVNSTINPAC-YALC	NATFKKTFRHLILCQRYNIGTAR
8	LMHIGYWLCYINSTVNPIC-YALC	NRTFRKTFKMLLLCRWKKKKVVEKLYWQGNKSLP
9	LFVFFNWLGYANSFNPFI-YCRS	PDFRKAFCQLCCARRARRRHHATHGDRPRASGCLARPGPSPGAASDDDDDDWGATPPARLLEPWAGCN- (25)
10	VYILLNWIGYVNSGFNPFI-YCRS	PDFRIAFQELLCLRRSSSLKAYGNGYSNGNTGEQSGYHVEQEKENKLICEDLPCTEDFVGHQGTVPSONID- (13)
11	AFALNWLGYANSFNPFI-YCRS	PDFRSARFRLLCRCGRRLPPEPCAAARPALFPGVPAEASSPAQPRLCQRLDG
12	VFKIAFWLGYNSCINPFI-YPCS	SQEFKKAQFQNLRIQCLRRKQSSKHTLCYTLHAPSHVLEGQHKDLVRI PVGSAETFYKI SKTDGVCWEKIF- (66)
13	VFKVFWLGYNSCINPFI-YPCS	SKEFKRAFMRJLCCQCRGRRRRRRRLGACATYTPWTRGCSLERSQSRKDSLDDSGSCMSGQKRTLP SA- (93)
14	LFKFFFWIGYCNSSNPFI-YTVF	NQDFRPFKHLIFRRRRRFRQ
15	LFQFFFWIGYCNSSNPFI-YTIF	NQDFRRFRRLILCRPWTQTAW
16	LFKFFFWIGYCNSSNPFI-YTIF	NHDFRRFRKILCRGDRKRIV
17	LFNFFFWIGYCNSSNPFI-YTIF	NHDFRRFRKILCRGDRKRIV
18	FNKFTWLGYNISNPFI-YTIF	NLDYRRFRKRLGILN
19	TFDVFVWFGWANSNPFI-YA-F	NADFRKAFSTLLGCYRLCPATNNAIETVSINNGAAMFSSHHEPRGSSISKECNLVYLIPHAVGSSSEDLKKE- (42)
20	TFDVFVWFGWANSNPFI-YA-F	NADFOKVAQLGCSHFCSRTPEVTNINELI SYNQDIVFHKEIAAAYIHMNPNAVTPGNREVONDEEG- (45)
21	LYSAFTWLGYNISNPFI-YTTF	NIEFRKAFKILHC
22	LYSATTWLGYNISNPFI-YTTF	NIEFRKAFKILSC
23	LYSAFTWLGYNISNPFI-YTVF	NAEFRNVFRKALRACC
24	LFDFFTWLGYNISNPFI-YTVF	NEEFRQAFQKIVPFRKAS
25	LGAIINWLGYNISNPFI-YAYF	NKDFQNAFKKIKCNFCRQ
26	LLNVFWIGYVCSGPNPFI-YTLF	NKIYRRAF SKYLRCQYKDPKPPVRIQIPRAATALSRELVNVIYRHTNERVARKANDPEPGIENOVENLE- (16)
27	LLNVFWIGYVCSGPNPFI-YTLF	NKTYRSASFRLQCYQYKRNKPLQILVMTIPALAYKSQQLVQGGKNSQEDAEQTVDDCSMTLKGQSE- (17)
28	LEATVWLGYNISNPFI-YAAL	NROFRITGYQLFCCRLANRNSHKTSLRNSASQLSRQTSREPRQEEKPLKLVMSGTEVTAPQGATR
29	AVDVTSAFAFFNSCINPFI-YVFM	GQDFRERLIHALPASLERALTEDSTOTSDTAINSTLPSAEVALQAK
30	LDLSCVSFAFFNSCINPFI-YVFA	GQGFQGRRLKSLPSLLRNVLTEESVVRKSFSTRSTVDTMAQKTOAV
31	AYLLCVCVSSISSCIDPLI-YYFA	SSECQRYVYSILCKEESDPPSSYNSSGQIMASKMDTCSNNIANSIYKLLT
32	-LLIYLRVATWNLDPFI-YILF	RRAVLRRLQPLRSLTRPSLSLQPLTQSRGLQ
33	ALDATEILGILHSCINPFI-YAFI	GQFRHGLKILAIHGLISKDSLPRKSRPSFVSSSGHTSTTL
34	AHQVTLCLLSTNCVDPFI-YCFL	TKKFRKHLSEKLNIMRSQKCSRVITDTGTETMAIPINHTFVNPFIKN
35	MYIGINLAIMNSCINPIALYFVS	KKFKNCFQSCLCCECYQSKSMTSVPMQTSIQWKNHEQNNHNTERSSHKDSIN
36	LDYIGINLAIMNSCINPIALYFVS	KRFKNCFKSCLCCECYQSKSMTSVPMQTSIQWKNHEQNNHNTERSSHKDSIN
37	TSICARLAPTNSCVNPFI-YLFS	KSEFQFNTQLCCQPGIMNRSHSTGRSTTCMTSFKSTNPSATFSLINRNI CHEGYV
38	VTIARLAPTNSCVNPFI-YLFS	ESFRKHSNQLCCGQKSYPERSTSYLLSSAVRMTSLKSNKNNVTVNSVLINGHSTKQEIAL
39	ALHVTQCLSLVHCCNPFI-YSFI	NRNRYEILMKAFIFKYSAKTGLTKLIDASRVSETEYSALQNAK
40	FYMLTALFVYSAINPFI-YNLV	SANFRQVFLSTLACLCGWRHRRKRPITSRKPNMSNNHAFSTSATRETL
41	VTQISSVAYNSCINPFI-YVIV	GKFRFKKSREYVQAI CRKGGCGESVQMSNGTLRTSISVDRQIHLQDWAGNQ
42	-LLKCRICILYNSAINPFI-YNLM	SQKFAAFRKLKNCQKQKTEKAAANSVALNYSVKESTRFSTELEDITVDTYVSTTKVSFDDTCLASEN
43	VYLALFWLAMSSTMYNPFI-YCCL	NHRFRSGFRLAFRCPPWPTTKEDKLELTPITSLSTRVNRCHTKETLFMAGDTAPSEATSGEAGRPQDGS- (17)
44	VYLASMWLMSSIMYNPFI-YCCL	NDRFRGLFKHAFRCPPFI SAGDYEGLEMKSTRYLQ- QSSVYKVSLETITSTVGAHEEPEEGPKATPS- (29)
45	VYLASFWMSSIMYNPFI-YCCL	NKRFRAGFKRAFRCPPFI QVSSYDELELKTTRFHPTRQSSLYTVSRMESVTVLFDPNQDPTKSSRKRRAV- (34)
46	AMPITICLAYFQNLNPLF-YGFL	GKKFKYFLQLLKYIPPAKSHSNLSTKMTLSYRPSQGNSSSTKPPAPCIEVE
47	LHHSILFSTINSSANPFI-YFFV	GSSKKRKFESLKVWLTRAFKDEMQRQKQKNC-NITVTVETV
48	VLLVLFYPI-NSCANPFI-YAIF	TKTFQRDFLLLSKFGCCRRADIYRRKDSAYTSNCKNGFTGSKNPSQSTLKLSTLHCQGTALIDKTRYEC
49	ILLVLFYPL-NSCANPFI-YAIF	TKAFQRDVFILLSKFGICRQQAAYRGQVRPPKNSDTIQVQKVTHDMRGALAMEDVVELIENSHLTPKKQ- (12)
50	ILLVLFHPI-NSCANPFI-YAIF	TKNFRDFFILLSKCCYEMQAIYRTETSSVTHNTHPRNGHCSAPRVTSGSSTYILVPLSLHQAQ
51	FMTIPAFFAKSAIYNPFI-YIMM	NKQFRNCMLTTICCGKNPLGDDEASATVSKTETSQVAPA
52	MAALPAFFAKSAIYNPFI-YVFM	NROFRNCILQFGKKVDDGSELSSASKTEVSVSVSPA
53	MAALPAYFAKSAIYNPFI-YVFM	NROFRNCILQFGKKVDDGSELSSASKTEVSVSVSPA
54	LVTIPSFFKSACIYNPFI-YCFM	NKQFQACIMQVCGKAMTDESDTCSQKTEVSVSVSQVGN
55	-TASVMTVTPMNPFI-YSL-	RNKDVKSVLKKTICEEIVRSPSLHFFLVLCPLCFICY
56	-AAAVMYAVTPMNPFI-YSL-	RNSDMKAAALRVKLMRFPKQ
57	-AITVLNTIVTPMNPFI-YTL-	RNKDVKEALRRTVKGK
58	-SASVMTVTPMNPFI-YSL-	RNKDVKRALERLLEGNCKVHHWTG
59	-VMAMMTVTPMNPFI-YSL-	RNRDMKRALIRVICSMKITL
60	-LVSVLAVIVPLNPFI-YCL-	RNQDVKRALRRLHLAQDQEAANTNGSKIG
61	-AMAMMTVTPMNPFI-YSL-	RNRDMKQALIRVTCSSKISLPW
62	-VMSLMTVTPMNPFI-YSL-	RNRDIKDALEKIMCKKQIPSL
63	-AMAMMTVTPMNPFI-YSL-	RNRDMKRALIRVICTKKISL
64	-VMAMMTVTPMNPFI-YSL-	RNRDMKEALIRVLCCKKITL
65	VFAFCSMLCLNSTVNPFI-YALR	SKDLRHAFRSMFPCSEGAQPLDMSGSDCLHKKHANTASMHRAAESICKSTVKIAKVMTSVSDTSAEAL
66	LYFAFWLWAMSTCYNPFI-YCWL	NENFRVELKALLSMCQRPPEDRLPSPVPSFRVAMTEKSHGRAPLPHHLPSSQIQSGKTDLSSEVPVWMS
67	LFLLCHLTAMISTCVNPIF-YGFL	NKNFQRLQFFNFCDFRSROGRITRL
68	AEYFLV-LAVINSGTNPFI-YTIL	NKEMRAFI RIMSCCKCPGDSACKFKRPIIAGMEFSRSKSDNSHPKQDEGNPETIMSSGNVNS
69	AMCLGILLSHANSMPNPIV-YACK	KKFKEYTVILRACRLCQTSDSLSDNLEQTE
70	VQSLSVILGYANSANPFI-YGFL	SDNFKRSFORILCLSMNDNAEPEVDYATALKSRAYSVEDFQENLESQGVFRNGTCASRISTL
71	PEYVTDLCICINSSAKPIV-YFLA	GRDKSQRLEPLRVVQRLRDAEPGDAASSTPMTVIMENQCPSGNAS
72	IGTLARVPHLCLINPFI-YALL	GHDFLQRMQCFRQGLDRAFLRSQONORA
73	CLYVGQFLAYVACINPFI-YILV	GTQMRKDMWTLRVFACCCVQKEIPIYQIDI
74	ALILTESLAFCHCLNPLF-YVVF	GTFRKKNYTVCPFSASDSFPAMYPGTTA

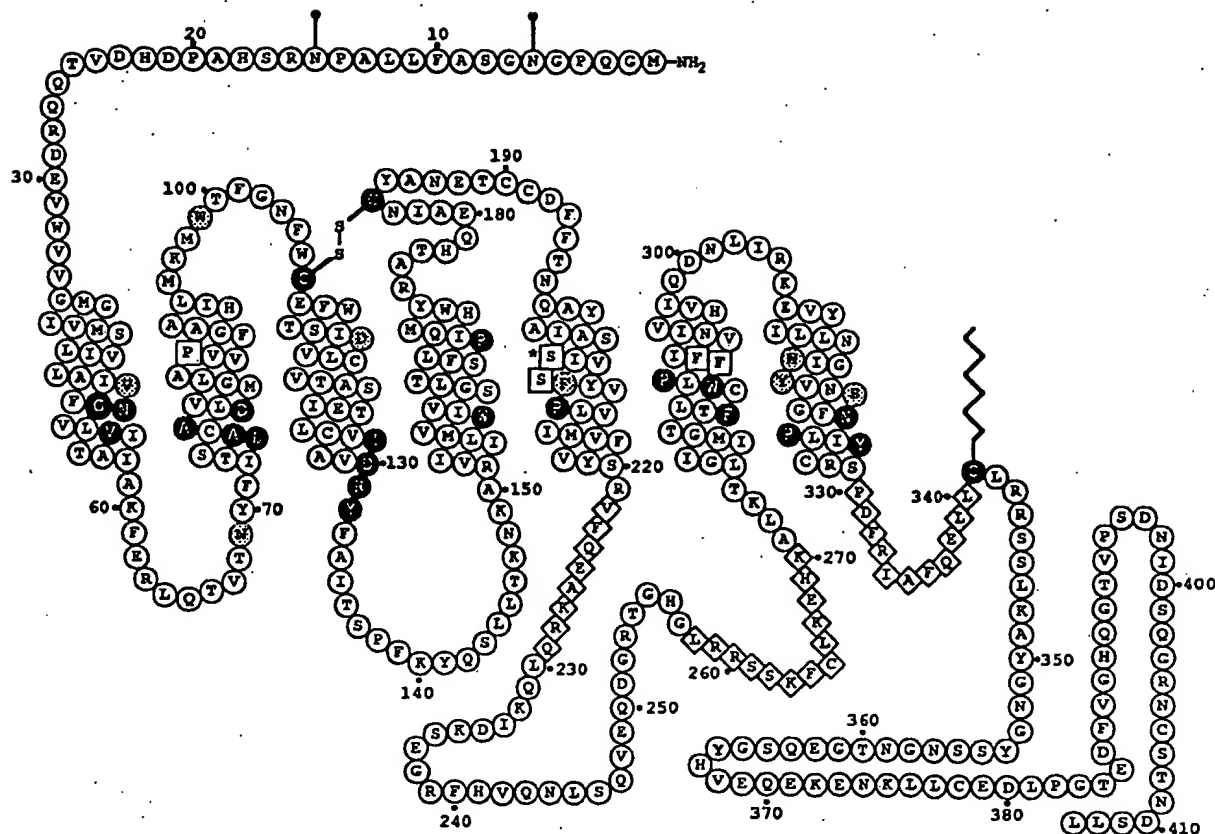


FIG. 3. Model of the human β_2 -adrenergic receptor. Amino acid residues in black are conserved in nearly all of the GPRs. Stippled residues are conserved in cationic amine receptors. Boxed residues are conserved in all catecholamine receptors. The asterisk denotes the serine conserved in the serotonin receptors. Residues in diamonds are residues believed to be involved in G-protein coupling. Glycosylated asparagines (N-6 and N-15) within the amino terminus are indicated. Cys³⁴¹ of the carboxyl terminus is known to be palmitoylated. Protein kinase A phosphorylation sites are indicated by arrowheads.

LIGAND BINDING DOMAINS

Our understanding of the structure of the binding site of the GPRs and of which residues actually interact with agonists and antagonists is rapidly evolving. The ligand binding sites of rhodopsin and of the adrenergic and muscarinic receptors have been partly delineated through biochemical and molecular biological approaches (for review, see Applebury and Hargrave, 1986; O'Dowd *et al.*, 1989b; Strader *et al.*, 1989b; Venter *et al.*, 1989; Hulme *et al.*, 1990). For most of the GPRs, with the possible exception of the glycoprotein hormone receptors, the ligand binding pocket appears to be created by the membrane-spanning regions.

As none of the GPRs have yet been crystallized, modeling of the three-dimensional array of the helices is based on the structure of bacteriorhodopsin, which has recently been resolved at high resolution (Henderson *et al.*, 1990). The transmembrane domains appear to form a hydrophilic pocket for ligand binding surrounded by hydrophobic residues (Strader *et al.*, 1989b; Venter *et al.*, 1989; Hulme *et al.*,

1990). The putative arrangement of the residues around the ligand binding site have been analyzed through helical wheel modeling. The α -helices contains 3.6 residues per helical turn. When the assortment of residues around the helix is predicted for the muscarinic receptors (Hulme *et al.*, 1990), adrenergic receptors (Strader *et al.*, 1989b; Venter *et al.*, 1989), and many other GPRs (Donnelly *et al.*, 1989), the domains contain a predominance of hydrophobic residues on one side and hydrophilic on the other. The hydrophilic side of each helix is postulated to face inwards and form the polar ligand binding site. Recently, computer-generated models for ligand-receptor interactions have been developed (Findlay and Eliopoulos, 1990; Henderson *et al.*, 1990; Dahl *et al.*, 1990; Hibert *et al.*, 1991).

The presence of a ligand binding pocket for the chromophore retinal deep within the transmembrane α -helices of rhodopsin was suggested by cross-linking and fluorescent energy transfer studies (Hargrave *et al.*, 1982; Thomas and Stryer, 1982). Retinal forms a Schiff base linkage with Lys²⁹⁶ in TM 7 (Thomas and Stryer, 1982). The Glu¹¹³ of

rhodopsin in TM 3 has been proposed as a counterion that interacts with the protonated Schiff base retinal, although mutagenesis studies have been inconclusive (Sakmar *et al.*, 1989; Zhukovsky and Oprian, 1989).

A variety of approaches support the existence of a similar intrahelical binding site in the cationic amine receptors. The ligand binding site of the adrenergic receptors has been investigated by photoaffinity labeling (Bar-Sinai *et al.*, 1986; Dohlman *et al.*, 1988), fluorescence emission spectroscopy (Tota and Strader, 1990), deletion mutants (Dixon *et al.*, 1987a,b), site-directed mutagenesis (Chung *et al.*, 1988; Dixon *et al.*, 1988; Strader *et al.*, 1988; Fraser, 1989; Strader *et al.*, 1989a,b; Wang *et al.*, 1991), and receptor chimeras (Kobilka *et al.*, 1988). As was the case for the visual pigments, the transmembrane domains are necessary for ligand binding and confer ligand specificity, while the hydrophilic extracellular and intracellular domains are not directly involved in ligand binding (Dixon *et al.*, 1981a,b). In both the α - and β -adrenergic receptors (Strader *et al.*, 1988; Wang *et al.*, 1991) as well as the m, muscarinic receptor (Fraser *et al.*, 1989), site-directed mutagenesis has demonstrated that the TM 3 aspartate (Asp¹¹³ in the β -adrenergic receptor; see Fig. 3) is critical for wild-type agonist and antagonist binding. The cationic amines, which include epinephrine, norepinephrine, dopamine, serotonin, and acetylcholine, all contain a positively charged amine head group which most likely interacts with the conserved TM 3 aspartate found in these receptors.

Mutagenesis studies have suggested that particular residues conserved within receptor subclasses can contribute to agonist specificity. Two conserved serines in TM 5 (Ser²⁰⁴ and Ser²⁰⁷ in the β -adrenergic receptor) have been implicated in forming hydrogen bonds with the *meta*- and *para*-hydroxyl groups of adrenergic agonists. Replacement of either serine by alanine reduces agonist binding to the same degree as removing the corresponding hydroxyl group from the ligand (Strader *et al.*, 1989a). Recent mutagenesis studies of the α_2 -adrenergic receptor suggest that Ser²⁰⁴ (corresponding in position to Ser²⁰⁷ of the β -adrenergic receptor) binds in an analogous fashion to the *para*-hydroxyl group of adrenergic ligands (Wang *et al.*, 1991). Two corresponding serine residues are found in TM 5 in all the dopamine receptors and a single conserved serine residue in TM 5 of the serotonin receptors (Fig. 2). The similarity of ligand structure and receptor sequence suggests that these TM 5 serines may also hydrogen bond with the aromatic hydroxyl groups of their respective agonists. This hypothesis is supported by computer modeling of the ligand-receptor interaction (Hibert *et al.*, 1991). The muscarinic receptors all contain a conserved asparagine in TM 6, not found in any other receptor subclass, which has been proposed to interact with the ester group of acetylcholine (Hibert *et al.*, 1991). Conserved TM 6 and/or TM 7 aromatic residues (e.g., Phe²⁹⁰ and Tyr³²⁰ in the β -adrenergic receptor) may interact with the aryl ring of serotonergic and adrenergic ligands (Dixon *et al.*, 1988; Hibert *et al.*, 1991).

The ligands for the glycoprotein hormone receptors, thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), and luteinizing hormone/chorionic

gonadotropin (LH/CG), are much larger than the ligands for the other GPRs. Presumably because of the large size of the ligands, this receptor subclass has evolved a distinct structure containing an extremely long first cytoplasmic domain encompassing the high-affinity hormone binding site. This glycosylated extracellular domain is rich in cysteine residues that may form disulfide bridges and help maintain the three-dimensional structure of the proteins (Sprengel *et al.*, 1990). The large amino-terminal extracellular domain of these receptors contains multiple leucine-rich repeats that identify these GPRs as members of a second gene family, that of the leucine-rich glycoprotein family (Takahashi *et al.*, 1985; Krusius and Ruoslahti, 1986). The extracellular location of a hormone binding site is supported by chimera studies (Moyle *et al.*, 1991; Nagayama *et al.*, 1991b) and, in the case of the LH/CG receptor, by the secretion of a soluble hormone binding protein generated by alternative splicing which encompasses only the amino terminus (Loosfelt *et al.*, 1989; Tsai-Morris *et al.*, 1990). Short regions of the amino terminus of TSH and LH/CG are necessary for high-affinity hormone binding (Wadsworth *et al.*, 1990; Nagayama *et al.*, 1991a,b). In contrast, β -adrenergic receptor ligand binding is not dependent on the amino-terminal extracellular domain (Dixon *et al.*, 1987b).

Recently the binding and activation of an LH/CG receptor construct in which virtually the entire extracellular amino terminus has been deleted was investigated (Ji and Ji, 1991b). The finding that CG can bind to the seven transmembrane components of the receptor, albeit with lower affinity, in the absence of the extracellular amino terminus suggests that this receptor may contain both a high-affinity binding site extracellularly and a low-affinity site within the transmembrane domains. CG binding to this low-affinity receptor mutant was capable of stimulating cAMP production. Possibly the high-affinity extracellular binding site serves to capture the hormone and present it to the intramembranous binding pocket for signal transduction.

INTRACELLULAR COUPLING

The GPRs are coupled by heterotrimeric G-proteins to various intracellular enzymes, ion channels, and transporters (Johnson and Dhanasekaran, 1989; Birnbaumer *et al.*, 1990). The G-proteins, which associate with GPRs at the intracellular face of the plasma membrane, are composed of relatively invariant β - and γ -subunits and a variable α -subunit (α_s , α_i , α_o) for which the G-protein is named (G_s , G_i , G_o). By a process not yet understood, GPR agonist binding catalyzes the exchange of GTP for GDP on the α -subunit (G-protein "activation"), resulting in its dissociation and stimulation of one (or more) of the various signal-transducing enzymes and channels. The different G-protein α -subunits preferentially stimulate particular effectors. The specificity of signal transduction may be determined, therefore, by the specificity of G-protein coupling. Some GPR residues or regions which are necessary for efficient signal transduction can be postulated to interact

with conserved G-protein motifs. In addition, certain short amino acid stretches of the receptors which are necessary for G-protein coupling also determine the specificity of the G-protein interactions.

Three types of studies investigating the relationship between receptor structure and G-protein affinity have been performed. Deletion and site-directed mutagenesis studies implicate receptor regions and amino acid residues that are necessary for efficient G-protein coupling. Synthetic peptide competition studies suggest which oligopeptide domains may directly interact with the G-proteins. Chimera experiments delineate the receptor regions that determine the specificity of G-protein coupling. Certain general principles arise from these multifaceted investigations. All of the intracellular domains are implicated in efficient G-protein coupling of various receptors. Short stretches of the membrane proximal regions of the third cytoplasmic loop and possibly the carboxyl terminus appear particularly critical in determining the specificity of G-protein coupling for many receptors.

Single residue mutations in the cytoplasmic loops of the β -adrenergic receptor reduce signal transduction (Dixon *et al.*, 1988; O'Dowd *et al.*, 1988). Site-directed mutagenesis of a conserved proline in the second intracellular loop to threonine, for example, caused no change in agonist binding but a ~35% reduction in adenylate cyclase stimulation (O'Dowd *et al.*, 1988). Site-directed mutagenesis has identified particular charged residues in the membrane proximal regions of the second and third intracellular loops which contribute to efficient G-protein coupling. Mutation of the highly conserved aspartate adjacent to TM 3 in the second intracellular loop of the β -adrenergic receptor (Asp¹³⁹ which is in the "DRY" sequence) gives rise to a receptor with high-affinity ligand binding but reduced or absent G-protein coupling (Dixon *et al.*, 1988; Fraser *et al.*, 1988). Similar results have been obtained for the muscarinic m₁ and the α_{2A} -adrenergic receptors (Fraser *et al.*, 1988, 1989; Wang *et al.*, 1991). The corresponding glutamate of rhodopsin is similarly implicated as interacting with transducin (Franke *et al.*, 1990). Another residue needed for transducin activation by rhodopsin is the lysine located in the distal third intracellular loop, Lys²⁴⁶. Mutation of this lysine to leucine results in a complete loss of signal transduction (Franke *et al.*, 1988). Mutation or deletion of histamine at the corresponding position in the β -adrenergic receptor reduced, although did not abolish, adenylate cyclase stimulation (O'Dowd *et al.*, 1988).

The TM 2 aspartate (Asp⁷⁹ of the β -adrenergic receptor, see Fig. 3), which is conserved in virtually all GPRs, is necessary for wild-type agonist binding and G-protein activation in many GPRs studied (Chung *et al.*, 1988; Strader *et al.*, 1988; Fraser *et al.*, 1989; Wang *et al.*, 1991; Ji and Ji, 1991a). In the α_1 -adrenergic and dopamine D₂ receptors, this aspartate is essential for modulation of receptor coupling by Na⁺ and H⁺, possibly due to allosteric modulation of receptor conformation (Horstman *et al.*, 1990; Neve, 1991; Neve *et al.*, 1991). Another transmembrane residue, the TM 6 cysteine found in most GPRs, has been implicated in β -adrenergic receptor signal transduction (Fraser, 1989).

Deletion studies of the β_2 -adrenergic receptor have indicated that the membrane proximal regions of the third cytoplasmic loop (residues 222-229 and 258-270, see Fig. 3) are necessary for signal transduction (Strader *et al.*, 1987a; O'Dowd *et al.*, 1988). In the α_1 -adrenergic receptor, deletion of seven amino acids of the third intracellular loop proximal to TM 7 caused a marked reduction in coupling to phospholipase C (Cotecchia *et al.*, 1990).

Deletions of carboxy-terminal residues adjacent to TM 7 produce mutant rhodopsin or β_2 -adrenergic receptors with diminished ability to activate G-proteins (O'Dowd *et al.*, 1988; Franke *et al.*, 1990). Mutation of a palmitoylated carboxy-terminal Cys³⁴¹ to glycine markedly reduced agonist stimulation of adenylate cyclase of the β_2 -adrenergic receptor (O'Dowd *et al.*, 1989a). The palmitoylated cysteine is predicted to anchor the carboxyl terminus to the membrane, producing a fourth cytoplasmic loop (see Fig. 3). Membrane anchorage may optimally position carboxy-terminal residues for G-protein interaction (Ovchinnikov *et al.*, 1988; O'Dowd *et al.*, 1989a). Regions of the carboxyl terminus and third cytoplasmic loop, adjacent to the transmembrane domains, may form clustered amphipathic α -helices (Strader *et al.*, 1987a; Higashijima *et al.*, 1988; Strader *et al.*, 1989b; Palm *et al.*, 1990). These helices, along with charged intracellular residues of the second and third intracellular loops (*i.e.*, DRY), may cooperatively interact to efficiently bind and activate G-proteins.

The activation of G-proteins by amphipathic α -helices is supported by experiments in which the G proteins G_i and G_o have been directly activated by mastoparan and other small peptides which form amphipathic α -helices at the inner surface of the cytoplasmic membrane (Higashijima *et al.*, 1988, 1990). Furthermore, direct activation of G_s has been demonstrated for synthetic peptides representing the third intracellular loop sequences adjacent to TM 5 and TM 6 of the β_2 -adrenergic receptor (Cheung *et al.*, 1991), and by a peptide representing the intracellular third loop sequence proximal to TM 6 of the avian β -adrenergic receptor (Palm *et al.*, 1989; Munch *et al.*, 1991).

Peptide competition experiments, in which short synthetic peptides competitively bind to G-proteins but do not activate them, have been useful in mapping GPR regions that are likely to contact the G-proteins. Receptor uncoupling following mutagenesis or deletion of receptor segments may be due either to loss of G-protein contacts or to altered tertiary structure of the receptors. Competition studies have been invaluable, therefore, in confirming that the loss of signal transduction observed in deletion and mutagenesis studies involves residues that directly bind the G-proteins. The regions of various receptors implicated by peptide competition studies include the membrane proximal regions of all three cytoplasmic loops and the carboxyl terminus of the avian β -adrenergic receptor (Palm *et al.*, 1989; Munch *et al.*, 1991), the second intracellular loop and the carboxyl terminus of the third intracellular loop of the α_{2A} -adrenergic receptor (Dalman and Neubig, 1991), and the second and third intracellular loops and amino-terminal region of the carboxyl terminus of rhodopsin (Konig *et al.*, 1989).

Chimera experiments involving hybrid α_2/β_2 -adrenergic

receptors suggested that the third cytoplasmic loop may underlie coupling specificity of the adrenergic receptors (Kobilka *et al.*, 1988). The β_2 -receptor is positively coupled to adenylate cyclase through G_s , whereas the α_1 -receptor is negatively coupled to this enzyme through G_i . β_2 -Adrenergic receptors were generated in which the third cytoplasmic loop was replaced by the third cytoplasmic loop of the α_1 -adrenergic receptor. Activation of this chimeric receptor, which still has β_2 pharmacology, caused inhibition instead of stimulation of adenylate cyclase (Kobilka *et al.*, 1988). Similar results have been obtained for other cationic amine receptor hybrids. The dopamine D_2 receptor is negatively coupled to adenylate cyclase, whereas the β_2 -adrenergic receptor is stimulatory to adenylate cyclase. Substitution of the third cytoplasmic loop of the phospholipase-coupled muscarinic m_1 receptor into the dopamine D_2 receptor and of the same region of the phospholipase-linked α_1 -adrenergic into the β_2 -adrenergic receptor caused the resultant chimeras to hydrolyze phosphatidylinositol and mobilize calcium (Cotecchia *et al.*, 1990; England *et al.*, 1991).

The receptor region of the third cytoplasmic loop most important in determining the specificity of signal transduction may differ between the muscarinic and adrenergic receptors. The signal transduction of α_2/β_2 -adrenergic receptor chimeras, in which short segments of the membrane proximal regions of the third intracellular loops and of the carboxyl terminus have been exchanged, indicated that the segment of the third intracellular loop adjacent to TM 6 is most important in adrenergic receptor coupling specificity (Liggett *et al.*, 1991). Substitution of multiple segments suggested that all of these domains may coordinately contribute to G-protein coupling (Liggett *et al.*, 1991). By contrast, interchange of seven amino acids from the third intracellular loop adjacent to TM 5 was sufficient to change the coupling specificity of a muscarinic m_1/m_2 chimera (Kubo *et al.*, 1988). In another series of experiments, substitution of nine amino acids from the amino terminus of this region of the third cytoplasmic loop of the muscarinic m_1 receptor into the m_2 receptor conferred a pattern of calcium release characteristic of m_1 receptor activation (Lechleiter *et al.*, 1991).

Phosphorylation of cytoplasmic residues has been identified as an important mechanism for the regulation of G-protein coupling of some GPRs. The third cytoplasmic loop and carboxyl terminus are rich in serine and threonine residues that are potential phosphorylation sites. After activation, both rhodopsin and the β_2 -adrenergic receptors are desensitized through the action of receptor kinases. The photoactivated form of rhodopsin is phosphorylated in the carboxyl terminus by a specific rhodopsin kinase (Hargrave *et al.*, 1982). This phosphorylation allows binding of the protein arrestin, which interferes with G-protein coupling to the opsin. A similar mechanism has been identified for the β_2 -adrenergic receptor, in which β -adrenergic receptor kinase (β ARK) phosphorylates the carboxyl terminus of the receptor. This leads to binding of a β -arrestin and functional uncoupling of the receptor. β ARK can also phosphorylate the third cytoplasmic loops of agonist stimulated m_4 muscarinic and α_2 -adrenergic receptors, as well as the carboxyl terminus of photoactivated rhodopsin

(Benovic *et al.*, 1986, 1987a; Kwatra *et al.*, 1989). Many receptors contain cytoplasmic consensus sequences for protein kinase A phosphorylation. In the case of the β -adrenergic receptors these sites play a role in receptor desensitization (Clark *et al.*, 1989). The TSH receptor, which does not contain consensus sequences for protein kinase A phosphorylation, does not demonstrate agonist-induced desensitization (Takasu *et al.*, 1978).

Structure/function modeling of the mechanism of G-protein activation by GPRs must also account for the recent identification of G-protein coupled receptors which are not members of the GPR gene family and of peptides that are capable of directly activating G-proteins. The secretin receptor, while distinct in sequence, is predicted to exhibit a seven-transmembrane domain structure (Ishihara *et al.*, 1991). Although the metabotropic glutamate receptor also manifests seven closely spaced hydrophobic domains, the hydrophobicity profile predicts an additional potential membrane spanning domain distant from the other seven (Masu *et al.*, 1991). The activation of heterotrimeric G-proteins has been implicated in the signal transduction of several membrane receptor tyrosine kinases. These related receptors, which bear no overall structural or sequence resemblance to the GPR family, include the insulin receptor, the insulin-like growth factor-II receptor, the epidermal growth factor receptor, and the colony stimulating factor-1 receptor encoded by the *c-fms* proto-oncogene (Imamura and Kufe, 1988; Nishimoto *et al.*, 1989; Luttrell *et al.*, 1990; Liang and Garrison, 1991). A 14-amino-acid segment of the insulin-like growth factor-II receptor, which bears striking resemblance in its charge distribution to the amphipathic protein mastoparan and to the membrane proximal regions of the third cytoplasmic loop of the GPRs, specifically activates the heterotrimeric G-protein, G_i (Okamoto *et al.*, 1990; Nishimoto *et al.*, 1991). The oncogenic activity of the *v-fps* protein, a cytosolic tyrosine kinase, may also involve activation of a heterotrimeric G-protein (Alexandropoulos *et al.*, 1991). GAP-43 is a growth cone protein that activates G_o . A decapeptide domain of GAP-43 which is homologous to the membrane proximal carboxyl terminus of many GPRs was found to be responsible for association with G_o (Strittmatter *et al.*, 1990). Amphiphilic neuropeptides, including substance P, ACTH, and bradykinins, can also activate G-proteins in a receptor-independent fashion (for review see Mousli *et al.*, 1990). Further delineation of the structural motifs that mediate G-protein coupling of these non-homologous receptors and peptides would be expected to illuminate the mechanisms of receptor/G-protein interaction in general.

GENE STRUCTURE AND EVOLUTION

Molecular cloning has revealed that a panoply of receptor subtypes exist for most of the classical neurotransmitters. As the basic transmitters developed very early phylogenetically (see Walker and Holden-Dye, 1989), the subsequent evolution of multiple receptor subtypes served the need for greater signaling specificity of progressively

more complex nervous systems. Nucleotide sequence analysis and analysis of gene structure may elucidate the time frame and mechanisms of subfamily and subtype evolution.

The remarkable conservation of the transmembrane domains of the GPR family proteins suggests that these genes may have evolved from a common precursor. A phylogenetic tree of the GPR family, generated by nucleotide sequence comparison, suggests that the opsins diverged from the catecholamines between 1 and 1.5 billion years ago (Yokoyama *et al.*, 1989). The age of the GPR gene family is independently suggested to be greater than 1 billion years by the isolation of a *Dictyostelium* chemoattractant receptor with structural and sequence homology with the GPRs (Klein *et al.*, 1988). While several seven-transmembrane yeast pheromone receptors have been identified (Hagen *et al.*, 1986; Marsh and Herskowitz, 1988), these proteins have little amino acid sequence homology with the GPRs. The evolutionary relationship, if any, between these yeast receptors, the seven-transmembrane bacteriorhodopsin, and the GPR superfamily remains to be determined.

Many of the GPR genes characterized to date are intronless. Several GPR genes, however, have introns within their coding regions. These include the opsins (Nathans and Hogness, 1984; Nathans *et al.*, 1986), the dopamine D₂, D₃, and D₄ receptors (Grandy *et al.*, 1989; Sokoloff *et al.*, 1990; Gandelman *et al.*, 1991; Van-Tol *et al.*, 1991), the substance P receptor (Hershey *et al.*, 1991), the substance K receptor (Gerard *et al.*, 1990), the luteinizing hormone receptor (Frazier *et al.*, 1990; Tsai-Morris *et al.*, 1990), and a *Drosophila* muscarinic receptor gene (Shapiro *et al.*, 1989). Introns have not been found within the coding regions of mammalian muscarinic receptor genes isolated to date. For the GPR genes with introns, the locations of introns near or within the seven transmembrane domains, are illustrated in Fig. 4. Introns tend to be positioned between TM domains.

Two mechanisms of gene evolution, gene duplication and retroposition, both appear to have played a role in generating the complex multiplicity of the GPR family. Genes containing introns, like those for the dopamine D₂, D₃, and D₄ receptors, most likely evolved from each other by gene duplication (Ohno, 1970). This is supported by the relative preservation of intron location among these receptors (see Fig. 4). There is also the preservation of an intron site (adjacent to the region encoding TM 3), between the dopamine and tachykinin receptor genes and of a different intron site (adjacent to the region encoding TM 7) between the tachykinin receptor and opsin genes, suggesting that gene duplication of a common precursor may have played a role in the evolution of these receptors. Most of the receptor genes are intronless, raising the possibility that one or more of these arose through reverse transcription of mRNA and incorporation into the genome (Brosius, 1991), an event referred to as retroposition. Gene duplication may have further amplified the number of these intronless genes.

Another potential mechanism for generating functionally distinct receptors is alternative processing of RNA primary transcript. Alternative splicing of a free-standing exon of the dopamine D₂ receptor gene gives rise to two receptor isoforms which differ in the incorporation or absence of a 29-amino-acid segment of the third cytoplasmic loop (Grandy *et al.*, 1989). Although the functional difference between the two isoforms remains to be elucidated, their biological importance is suggested by the preservation of the alternative splice site through at least 80 million years of evolution, from mouse to man (Montmayeur *et al.*, 1991). Alternative splicing also gives rise to multiple forms of the dopamine D₂ receptor mRNA (Giros *et al.*, 1991; Snyder *et al.*, 1991) and of the LH/CG receptor. An alternative mRNA splice variant of the LH/CG receptor encodes a secreted LH binding protein lacking the transmembrane regions (Loosfelt *et al.*, 1989; Tsai-Morris *et al.*, 1990).

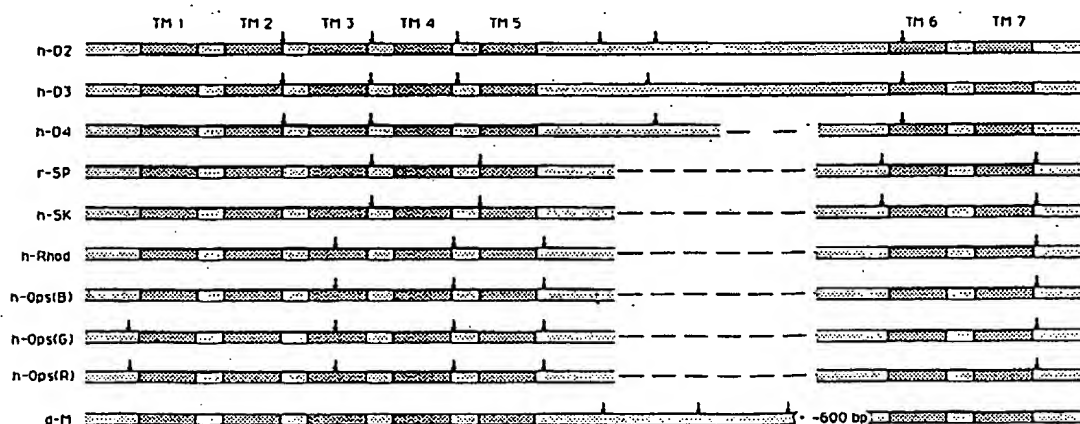


FIG. 4. Schematic representation of GPR genes that have introns within the protein coding region. Abbreviations: h-D₂, human dopamine D₂ receptor; r-D₂, rat dopamine D₂ receptor; h-D₄, human dopamine D₄ receptor; r-SP, rat substance P receptor; h-SK, human substance K receptor; h-Rhod, human rhodopsin; h-ops(B), human blue opsin; h-ops(G), human green opsin; h-ops(R), human red opsin; d-M, *Drosophila* muscarinic receptor. The locations of introns are indicated by arrows.

Convergent evolution is also evident in G-protein coupled receptors. The examples of secretin and the metabotropic glutamate receptor in which apparently unrelated genes have evolved similar seven-transmembrane structures and G-protein coupling have already been discussed. Comparison of the nucleotide sequence for the red and green visual pigment genes between fish and human indicate that the red pigments evolved independently from the green pigment through identical amino acid substitutions (Yokoyama and Yokoyama, 1990).

SUMMARY

The identification of new GPR genes and the elucidation of their binding and G-protein coupling mechanisms will undoubtedly continue to accelerate. In particular, the binding site and coupling domains of the non-glycoprotein hormone peptide receptors have not yet been investigated and their study will help illuminate the binding and coupling characteristics in this family. Although striking progress has been made in delineating the ligand binding site of rhodopsin and the cationic amine GPRs and the structural motifs contributing to G-protein recognition, the actual molecular events which transmit ligand binding into activation of G-protein remain to be elucidated. More complete understanding of the GPR's three-dimensional structure, pharmacology, physiology, and anatomy will ultimately have a tremendous impact on our understanding of biology and on the development of pharmaceuticals.

ACKNOWLEDGMENTS

We wish to thank Dr. James L. Roberts for advice and encouragement and Dr. Manami Tsutsumi for critically reading the manuscript. Supported by National Institutes of Health grants K11 DK01854 to S.C.S., MH 38819 to J.B., and MH 45212 to S.C.S. and J.B.

REFERENCES

- ALEXANDROPOULUS, K., JOSEPH, C.K., SPANGLER, R., and FOSTER, D.A. (1991). Evidence that a G-protein transduces signals initiated by the protein-tyrosine kinase v-Fps. *J. Biol. Chem.* 266, 15583-15586.
- APPLEBURY, M.L., and HARGRAVE, P.A. (1986). Molecular biology of the visual pigments. *Vision Res.* 26, 1881-1895.
- ARAI, H., HORI, S., ARAMORI, I., OHKUBO, H., and NAKANISHI, S. (1990). Cloning and expression of a cDNA encoding an endothelin receptor. *Nature* 348, 730-732.
- ARAKAWA, S., GOCAYNE, J.D., MCCOMBIE, W.R., URQUHART, D.A., HALL, F.C.M., and VENTER, J.C. (1990). Cloning, localization, and permanent expression of a *Drosophila* octopamine receptor. *Neuron* 4, 343-354.
- BAR-SINAI, A., ALDOUBY, Y., CHOREV, M., and LEVITSKY, A. (1986). Association of turkey erythrocyte β -adrenoceptors with a specific lipid component. *EMBO J.* 5, 1175-1180.
- BENOVIC, J.L., MAYOR, F.J., SOMERS, R.L., CARON, M.G., and LEFKOWITZ, R. (1986). Light-dependent phosphorylation of rhodopsin by beta-adrenergic receptor kinase. *Nature* 321, 869-872.
- BENOVIC, J.L., REGAN, J.W., MATSUI, H., MAYOR, F.J., COTECCHIA, S., LEEB, L.L.M., CARON, M.G., and LEFKOWITZ, R.J. (1987a). Agonist-dependent phosphorylation of the alpha 2-adrenergic by the beta-adrenergic receptor kinase. *J. Biol. Chem.* 262, 17251-17253.
- BENOVIC, J.L., STANISZEWSKI, C., CERIONE, R.A., CODINA, J., LEFKOWITZ, R.J., and CARON, M.G. (1987b). The mammalian beta-adrenergic receptor: Structural and characterization of the carbohydrate moiety. *J. Recept. Res.* 7, 257-281.
- BIRNBAUMER, L., ABRAMOWITZ, J., and BROWN, A.M. (1990). Receptor effector coupling by G-proteins. *Biochim. Biophys. Acta* 1031, 163-224.
- BONNER, T.I., YOUNG, A.C., BRANN, M.R., and BUCKLEY, N.J. (1988). Cloning and expression of the human and rat m5 muscarinic acetylcholine receptor genes. *Neuron* 1, 403-410.
- BOULAY, F., TARDIF, M., BROUCHON, L., and VIGNAIS, P. (1990). Synthesis and use of a novel N-formyl peptide derivative to isolate a human N-formyl peptide receptor cDNA. *Biochem. Biophys. Res. Commun.* 168, 1103-1109.
- BROSIOUS, J. (1991). Retroposons—Seeds of evolution. *Science* 251, 753.
- BUCK, L., and AXEL, R. (1991). A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* 65, 178-187.
- BUNZOW, J.R., VAN TOL, H.H., GRANDY, D.K., ALPERT, P., SALON, C.M., MACHIDA, C.A., NEVE, K.A., and CIVELLI, O. (1988). Cloning and expression of a rat D2 dopamine receptor cDNA. *Nature* 336, 783-787.
- CHEE, M.S., SATCHWELL, S.C., PREDDIE, E., WESTON, K.M., and BARRELL, B.G. (1990). Human cytomegalovirus encodes three G protein-coupled receptor homologues. *Nature* 344, 774-777.
- CHEUNG, A.H., HUANG, R.R., GRAZIANO, M.P., and STRADER, C.D. (1991). Specific activation of G_i by synthetic peptides corresponding to an intracellular loop of the beta-adrenergic receptor. *FEBS Lett.* 279, 277-280.
- CHUNG, F.Z., WANG, C.D., POTTER, P.C., VENTER, J.C., and FRASER, C.M. (1988). Site-directed mutagenesis and continuous expression of human β -adrenergic receptors. Identification of a conserved aspartate residue involved in agonist binding and receptor activation. *J. Biol. Chem.* 263, 4052-4055.
- CLARK, R.B., FRIEDMAN, J., DIXON, R.A.F., and STRADER, C.D. (1989). Identification of a specific site required for rapid desensitization of the beta-adrenergic receptor by cAMP-protein kinase. *Mol. Pharmacol.* 36, 343-348.
- COTECCHIA, S., EXUM, S., CARON, M.G., and LEFKOWITZ, R.J. (1990). Regions of the α_1 -adrenergic receptor involved in coupling to phosphatidylinositol hydrolysis and enhanced sensitivity of biological function. *Proc. Natl. Acad. Sci. USA* 87, 2896-2900.
- DAHL, S.G., EDVARDSEN, O., and SYLTE, I. (1991). Molecular dynamics of dopamine at the D_2 receptor. *Proc. Natl. Acad. Sci. USA* 88, 8111-8115.
- DALMAN, H.M., and NEUBIG, R.R. (1991). Two peptides from the α_2 -adrenergic receptor alter receptor G protein coupling by distinct mechanisms. *J. Biol. Chem.* 266, 11025-11019.
- DEARRY, A., GINGRICH, J.A., FALARDEAU, P., FREMEAUX, R.T.J., and CARON, M.G. (1990). Molecular cloning and expression of the gene for a human D1 dopamine receptor. *Nature* 347, 72-76.
- DIXON, R.A., KOBILKA, B.K., STRADER, D.J., BENOVIC,

- J.L., DOHLMAN, H.G., FRIELLE, T., BOLANOWSKI, M.A., BENNETT, C.D., RANDS, E., DIEHL, R.E., MUMFORD, R.A., SLATER, E.E., SIGAL, I.S., CARON, M.G., LEFKOWITZ, R.J., and STRADER, C.D. (1986). Cloning of the gene and cDNA for mammalian beta-adrenergic and homology with rhodopsin. *Nature* 321, 75-79.
- DIXON, R.A., SIGAL, I.S., CANDELORE, M.R., REGISTER, R.B., RANDS, E., and STRADER, C.D. (1987a). Structural features required for ligand binding to the beta-adrenergic receptor. *EMBO J.* 6, 3269-3275.
- DIXON, R.A., SIGAL, I.S., RANDS, E., REGISTER, R.B., CANDELORE, M.R., BLAKE, A.D., and STRADER, C.D. (1987b). Ligand binding to the beta-adrenergic receptor involves its rhodopsin-like core. *Nature* 326, 73-77.
- DIXON, R.A.F., SIGAL, I.S., and STRADER, C.D. (1988). Structure-function analysis of the beta-adrenergic receptor. *Cold Spring Harbor Symp. Quant. Biol.* 53, 487-498.
- DOHLMAN, H.G., CARON, M.G., and LEFKOWITZ, R.J. (1987). A family of receptors coupled to guanine nucleotide regulatory proteins. *Biochemistry* 26, 2657-2664.
- DOHLMAN, H.G., CARON, M.G., STRADER, C.D., AMLAIKY, N., and LEFKOWITZ, R.J. (1988). Identification and sequence of a binding site peptide of the beta2-adrenergic receptor. *Biochemistry* 21, 1813-1817.
- DONNELLY, D., JOHNSON, M.S., BLUNDELL, T.L., and SAUNDERS, J. (1989). An analysis of the periodicity of conserved residues in sequence alignments of G-protein coupled receptors. *FEBS Lett.* 251, 109-116.
- EMORINE, L.J., MARULLO, S., BRIEND, S.M.M., PATEY, G., DELAVIER, K.C., and STROBERG, A.D. (1989). Molecular characterization of the human beta 3-adrenergic. *Science* 245, 1118-1121.
- ENGLAND, B.P., ACKERMAN, M.S., and BARRETT, R.W. (1991). A chimeric D2 dopamine/m1 muscarinic receptor with D2 binding specificity mobilizes intracellular calcium in response to dopamine. *FEBS Lett.* 279, 87-90.
- ENGELMAN, D.M., HENDERSON, R., McLACHLAN, A.D., and WALLACE, B.A. (1980). Path of the polypeptide in bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA* 77, 2023-2027.
- EVA, C., KEINANEN, K., MONYER, H., SEEBURG, P., and SPRENGEL, R. (1990). Molecular cloning of a novel G protein-coupled receptor that may belong to the neuropeptide receptor family. *FEBS Lett.* 271, 81-84.
- FINDLAY, J., and ELIOPOULOS, E. (1990). Three-dimensional modelling of G protein-linked receptors. *Trends Pharmacol. Sci.* 11, 492-499.
- FRANKE, R.R., SAKMAR, T.P., OPRIAN, D.A., and KHORANA, H.G. (1988). A single amino acid substitution in rhodopsin (lysine 248-leucine) prevents activation of transducin. *J. Biol. Chem.* 263, 2119-2122.
- FRANKE, R.R., KONIG, B., SAKMAR, T.P., KHORANA, H.G., and HOFMANN, K.P. (1990). Rhodopsin mutants that bind but fail to activate transducin. *Science* 250, 123-125.
- FRASER, C.M. (1989). Site-directed mutagenesis of beta-adrenergic receptors: Identification of conserved cysteine residues that independently affect ligand binding and receptor activation. *J. Biol. Chem.* 264, 9266-9270.
- FRASER, C.M., CHUNG, F.Z., WANG, C.D., and VENTER, J.C. (1988). Site-directed mutagenesis of human beta-adrenergic receptors: Substitution of aspartic acid-130 by asparagine produces a receptor with high-affinity agonist binding that is uncoupled from adenylate cyclase. *Proc. Natl. Acad. Sci. USA* 85, 5478-5482.
- FRASER, C.M., WANG, C.D., ROBINSON, D.A., GOCAYNE, J.D., and VENTER, J.C. (1989). Site-directed mutagenesis of m1 muscarinic acetylcholine conserved aspartic acids play important roles in receptor function. *Mol. Pharmacol.* 36, 840-847.
- FRAZIER, A.L., ROBBINS, L.S., STORK, P.J., SPRENGEL, R., SEGALOFF, D.L., and CONE, R.D. (1990). Isolation of TSH and LH/CG receptor cDNAs from human thyroid: Regulation by tissue specific splicing. *Mol. Endocrinol.* 4, 1264-1276.
- FRIELLE, T., COLLINS, S., DANIEL, K.W., CARON, M.G., LEFKOWITZ, R.J., and KOBILKA, B.K. (1987). Cloning of the cDNA for the human beta 1-adrenergic receptor. *Proc. Natl. Acad. Sci. USA* 84, 7920-7924.
- GANDELMAN, K.Y., HARMON, S., TODD, R.D., and O'MALLEY, K.L. (1991). Analysis of the structure and expression of the human dopamine D2A receptor gene. *J. Neurochem.* 56, 1024-1029.
- GANTZ, I., MUNZERT, G., TASHIRO, T., SCHAFER, M., WANG, L., DELVALLE, J., and YAMADA, T. (1991). Molecular cloning of the human histamine H2 receptor. *Biochem. Biophys. Res. Commun.* 178, 1386-1392.
- GERARD, N.P., and GERARD, C. (1991). The chemotactic receptor for human C5a anaphylatoxin. *Nature* 349, 613-617.
- GERARD, N.P., EDDY, R.L., SHOWS, T.B., and GERARD, C. (1990). The neurokinin A (substance K) receptor. *J. Biol. Chem.* 265, 20455-20462.
- GIROS, B., MARTRES, M.-P., SOKOLOFF, P., and SCHWARTZ, J.-C. (1990). Clonage du gene du recepteur dopaminergique D3 humain et identification de son chromosome. *C.R. Acad. Sci. Paris* 311, 501-508.
- GIROS, B., MARTRES, M.P., PILON, C., SOKOLOFF, P., and SCHWARTZ, J.C. (1991). Shorter variants of the D3 dopamine receptor produced through various patterns of alternative splicing. *Biochem. Biophys. Res. Commun.* 176, 1594-1592.
- GRANDY, D.K., MARCHIONNI, M.A., MAKAM, H., STOFKO, R.E., ALFANO, M., FROTHINGHAM, L., FISCHER, J.B., BURKE, H.K., BUNZOW, J.R., SERVER, A.C., and CIVELLI, O. (1989). Cloning of the cDNA and gene for a human D2 dopamine receptor. *Proc. Natl. Acad. Sci. USA* 86, 9762-9766.
- HAGEN, D.C., MCCAFFREY, G., and SPRAGUE, G.F. (1986). Evidence the yeast STE3 gene encodes a receptor for the peptide pheromone a factor: Gene sequence and implications for the structure of the presumed receptor. *Proc. Natl. Acad. Sci. USA* 83, 1418-1422.
- HAMBLIN, M.W., and METCALF, M.A. (1991). Primary structure and functional characterization of a human 5-HT1D-type serotonin receptor. *Mol. Pharmacol.* 40, 143-148.
- HARGRAVE, P.A. (1977). The amino-terminal tryptic peptide of bovine rhodopsin. A glycopeptide containing two sites of oligosaccharide attachment. *Biochim. Biophys. Acta* 492, 83-94.
- HARGRAVE, P.A., MCDOWELL, J.H., SIEMIATKOWSKI, J.E.C., KUHN, H., WANG, J.K., CURTIS, D.R., MOHANA, R.J.K., ARGOS, P., and FELDMANN, R.J. (1982). The carboxyl-terminal one-third of bovine rhodopsin: Its structure and function. *Vision Res.* 22, 1429-1438.
- HARGRAVE, P.A., MCDOWELL, J.H., CURTIS, D.R., WANG, J.K., JUSZCZAK, E., FONG, S.L., RAO, J.K., and ARGOS, P. (1983). The structure of bovine rhodopsin. *Biophys. Struct. Mech.* 9, 235-244.
- HARRIGAN, M.T., CAMPBELL, N.F., and BOURGEOIS, S. (1991). Identification of a gene induced by glucocorticoids in murine T-cells: A potential G-protein coupled receptor. *Mol. Endocrinol.* 5, 1331-1338.
- HENDERSON, R., BALDWIN, J., CESKA, T.H., ZEMLIN, F., BECKMANN, E., and DOWNING, K. (1990). Model for

- the structure of bacteriorhodopsin based on high resolution electron cryomicroscopy. *J. Mol. Biol.* 213, 899-929.
- HERSHEY, A.D., DYKEMA, P.E., and KRAUSE, J.E. (1991). Organization, structure, and expression of the gene encoding the rat substance P receptor. *J. Biol. Chem.* 266, 4366-4374.
- HIBERT, M.F., TRUMPP-KALLMEYER, S., BRUINVELS, A., and HOFLACK, J. (1991). Three-dimensional models of neurotransmitter G-binding protein-coupled receptors. *Mol. Pharmacol.* 40, 8-15.
- HIGASHIJIMA, T., UZU, S., NAKAJIMA, T., and ROSS, E.M. (1988). Mastoparan, a peptide from wasp venom, mimics receptors by activating GTP-binding regulatory proteins (G-proteins). *J. Biol. Chem.* 262, 6491-6494.
- HIGASHIJIMA, T., BURNIER, J., and ROSS, E.M. (1990). Regulation of G_i and G_o by mastoparan, related amphiphilic peptides, and hydrophobic amines. *J. Biol. Chem.* 265, 14176-14186.
- HIRATA, M., HAYASHI, Y., USHIKUBI, F., TOKOTA, Y., KAGEYAMA, R., NAKANISHI, S., and NARUMIYA, S. (1991). Cloning and expression of cDNA for a human thromboxane A_2 receptor. *Nature* 349, 617-620.
- HLA, T., and MACIAG, T. (1990). An abundant transcript induced in differentiating human endothelial cells encodes a polypeptide with structural similarities to G-protein-coupled receptors. *J. Biol. Chem.* 265, 9308-9313.
- HONDA, Z., NAKAMURA, M., MIKI, I., MINAMI, M., WATANABI, T., SEYAMA, Y., OKADO, H., TOH, H., MIYAMOTO, T., and SHIMIZU, T. (1991). Cloning by functional expression of platelet-activating factor receptor from guinea-pig lung. *Nature* 349, 342-346.
- HORSTMAN, D.A., BRTANDON, S., WILSON, A.L., GUYER, C.A., CRAGOE, E.J., and LIMBIRD, L.W. (1990). An aspartate conserved among G-protein receptors confers allosteric regulation of α_2 -adrenergic receptors by sodium. *J. Biol. Chem.* 265, 21590-21595.
- HULME, E.C., BIRDSALL, N.J., and BUCKLEY, N.J. (1990). Muscarinic receptor subtypes. *Annu. Rev. Pharmacol. Toxicol.* 30, 633-673.
- IMAMURA, K., and KUFÉ, D. (1988). Colony-stimulating factor 1-induced Na^+ influx into human monocytes involves activation of a pertussis toxin-sensitive GTP-binding protein. *J. Biol. Chem.* 263, 14093-14098.
- ISHIHARA, T., NAKAMURA, S., KAZIRO, Y., TAKAHASHI, T., TAKAHASHI, K., and NAGATA, S. (1991). Molecular cloning and expression of a cDNA encoding the secretin receptor. *EMBO J.* 10, 1635-1641.
- JANSSEN, J.J.M., DECALUWE, G.L.J., and DEGRIP, W.J. (1990). Direct activation of GTP-binding regulatory proteins (G-proteins) by substance P and compound 48/80. *FEBS Lett.* 260, 113-118.
- JI, I., and JI, T.H. (1991a). Asp³⁰² in the second transmembrane domain of the lutropin receptor is important for high affinity hormone binding and cAMP production. *J. Biol. Chem.* 266, 14953-14957.
- JI, I., and JI, T.H. (1991b). Human choriogonadotropin binds to a lutropin receptor with essentially no N-terminal extension and stimulates cAMP synthesis. *J. Biol. Chem.* 266, 13076-13079.
- JOHNSON, G.J., and DHANASEKARAN, N. (1989). The G-protein family and their interaction with receptors. *Endoc. Rev.* 10, 317-331.
- JULIUS, D., HUANG, K.N., LIVELLI, T.J., AXEL, R., and JESSELL, T.M. (1990). The 5HT₂ receptor defines a family of structurally distinct but functionally conserved serotonin receptors. *Proc. Natl. Acad. Sci. USA* 87, 928-932.
- JULIUS, D., MACDERMOTT, A.B., AXEL, R., and JESSELL, T.M. (1988). Molecular characterization of a functional cDNA encoding the serotonin 1c receptor. *Science* 241, 558-564.
- KARNIK, S.S., SAKMANN, J.P., CHEN, H.B., and KHORANA, H.G. (1988). Cysteine residues 110 and 187 are essential for the formation of correct structure in bovine rhodopsin. *Proc. Natl. Acad. Sci. USA* 85, 8459-8463.
- KLEIN, P.S., SUN, T.J., SAXE, C.L., KIMMEL, A.R., JOHNSON, R.L., and DEVREOTES, P.N. (1988). A chemoattractant receptor controls development in *Dictyostelium discoideum*. *Science* 241, 1467-1472.
- KOBILKA, B.K., DIXON, R.A., FRIELLE, T., DOHLMAN, H.G., BOLANOWSKI, S.I.S., YANG, F.T.L., FRANCKE, U., CARON, M.G., and LEFKOWITZ, R.J. (1987a). cDNA for the human beta 2-adrenergic receptor: A protein with multiple membrane-spanning domains and encoded by a gene chromosomal location is shared with that of the receptor for platelet-derived growth factor. *Proc. Natl. Acad. Sci. USA* 84, 46-50.
- KOBILKA, B.K., FRIELLE, T., COLLINS, S., YANG, F.T., KOBILKA, T.S., FRANCKE, U., LEFKOWITZ, R.J., and CARON, M.G. (1987b). An intronless gene encoding a potential member of the family of receptors coupled to guanine nucleotide regulatory proteins. *Nature* 329, 75-79.
- KOBILKA, B.K., MATSUI, H., KOBILKA, T.S., YANG, F.T.L., CARON, M.G., LEFKOWITZ, R.J., and REGAN, J.W. (1987c). Cloning, sequencing, and expression of the gene coding for platelet alpha 2-adrenergic receptor. *Science* 238, 650-656.
- KOBILKA, B.K., KOBILKA, T.S., DANIEL, K., REGAN, J.W., and CARON, L.R.J. (1988). Chimeric alpha 2-beta 2-adrenergic receptors: Delineation of domains involved in effector coupling and ligand binding. *Science* 240, 1310-1316.
- KORNFELD, R., and KORNFELD, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 54, 631-664.
- KONIG, B., ARENDT, A., MCDOWELL, J.H., KAHLERT, M., HARGRAVE, P.A., and HOFMANN, K.P. (1989). Three cytoplasmic loops of rhodopsin interact with transducin. *Proc. Natl. Acad. Sci. USA* 86, 6878-6882.
- KRUSIUS, T., and RUOSLAHTI, E. (1986). Primary structure of an extracellular matrix proteoglycan core protein deduced from cloned cDNA. *Proc. Natl. Acad. Sci. USA* 83, 7683-7687.
- KUBO, T., BUJO, H., AKIBA, I., NAKAI, J., MISHINA, M., and NUMA, S. (1988). Location of a region of the muscarinic acetylcholine receptor involved in selective effector coupling. *FEBS Lett.* 241, 119-125.
- KWATRA, M.M., BENOVIĆ, J.L., CARON, M.G., LEFKOWITZ, R.J., and HOSEY, M.M. (1989). Phosphorylation of chick heart muscarinic cholinergic receptors by the beta-adrenergic receptor kinase. *Biochemistry* 28, 4543-4547.
- LANIER, S.M., DOWNING, S., DUZIC, E., and HOMCY, C.J. (1991). Isolation of rat genomic clones encoding subtypes of the alpha 2-adrenergic receptor. Identification of a unique receptor subtype. *J. Biol. Chem.* 266, 10470-10478.
- LECHLEITER, J., GIRARD, S., CLAPHAM, D., and PERALTA, E. (1991). Subcellular patterns of calcium release determined by G protein-specific residues of muscarinic receptors. *Nature* 350, 505-508.
- LIANG, M., and GARRISON, J.C. (1991). The epidermal growth factor is coupled to a pertussis toxin-sensitive guanine nucleotide regulatory protein in rat hepatocytes. *J. Biol. Chem.* 266, 13342-13349.
- LIBERT, F., LEFORT, A., GERARD, C., PARMENTIER, M., PERRET, J., LUDGATE, M., DUMONT, J.E., and VAS-

- SART, G. (1989a). Cloning, sequencing and expression of the human thyrotropin (TSH) receptor: Evidence for binding auto-antibodies. *Biochem. Biophys. Res. Commun.* 165, 1250-1255.
- LIBERT, F., PARMENTIER, M., LEFORT, A., DINSART, C., VAN, S.J., MAENHAUT, C., SIMONS, M.J., DUMONT, J.E., and VASSART, G. (1989b). Selective amplification and cloning of four new members of the G protein-coupled receptor family. *Science* 244, 569-572.
- LIGGETT, S.B., CARON, M.G., LEFKOWITZ, R.J., and HNATOWICH, M. (1991). Coupling of a mutated form of the human β_2 -adrenergic receptor to Gi and Gs: Requirement for multiple cytoplasmic domains in the coupling process. *J. Biol. Chem.* 266, 4816-4821.
- LILES, W.C., and NATHANSON, N.M. (1986). Regulation of neuronal muscarinic receptor number by protein glycosylation. *J. Neurochem.* 46, 85-95.
- LOMASNEY, J.W., LORENZ, W., ALLEN, L.F., KING, K., REGAN, J.W., YANG, F.T.L., CARON, M.G., and LEFKOWITZ, R.J. (1990). Expansion of the alpha 2-adrenergic receptor family: Characterization of a human alpha 2-adrenergic receptor gene for which is located on chromosome 2. *Proc. Natl. Acad. Sci. USA* 87, 5094-5098.
- LOOSFELT, H., MISRAHI, M., ATGER, M., SALESSE, R., THI, M., JOLIVET, A., GUIOCHON-MANTEL, A., SAR, S., JALLAL, B., GARNIER, J., and MILGROM, E. (1989). Cloning and sequencing of porcine LH-hCG receptor cDNA: Variants lacking transmembrane domain. *Science* 245, 525-528.
- LUTTREL, L., KILGOUR, E., LARNER, J., and ROMERO, G. (1990). A pertussis toxin-sensitive G-protein mediates some aspects of insulin action in BC3H-1 murine myocytes. *J. Biol. Chem.* 265, 16873-16879.
- MARSH, L., and HERSKOWITZ, I. (1988). STE2 protein of *Saccaromyces kluyveri* is a member of the rhodopsin/ β -adrenergic receptor family and is responsible for recognition of the peptide ligand α factor. *Proc. Natl. Acad. Sci. USA* 85, 3855-3859.
- MASU, M., TANABE, Y., TSUCHIDA, K., SHIGEMOTO, R., and NAKANISHI, S. (1991). Sequence and expression of a metabotropic glutamate receptor. *Nature* 349, 760-765.
- MATSUDA, L.A., LOLAIT, S.J., BROWNSTEIN, M.J., YOUNG, A.C., and BONNER (1990). Structure of a cannabinoid receptor and functional expression of cloned cDNA. *Nature* 346, 561-564.
- MCEACHERN, A.E., SHELTON, E.R., BHAKTA, S., OBERNOLTE, R., BACH, C., ZUPPAN, P., FUJISAKI, J., ALDRICH, R.W., and JARNAGIN, K. (1991). Expression cloning of a rat B2 bradykinin receptor. *Proc. Natl. Acad. Sci. USA* 88, 7724-7728.
- MEYERHOF, W., MÜLLER-BRECHLIN, R., and RICHTER, D. (1991a). Molecular cloning of a novel putative G-protein coupled receptor expressed during rat spermiogenesis. *FEBS Lett.* 284, 155-160.
- MEYERHOF, W., PAUST, H.-J., SCHONROCK, C., and RICHTER, D. (1991b). Cloning of a cDNA encoding a novel putative G-protein coupled receptor expressed in specific rat brain regions. *DNA Cell Biol.* 10, 689-694.
- MINEGISHI, T., NAKAMURA, K., TAKAKURA, Y., IBUKI, Y., and IGARASHI, M. (1991). Cloning and sequencing of human FSH receptor cDNA. *Biochem. Biophys. Res. Commun.* 175, 1125-1130.
- MONTMAYEUR, J.P., BAUSERO, P., AMLAIKY, N., MAROTEXUX, L., HEN, R., and BORRELLI, E. (1991). Differential expression of the mouse D2 dopamine receptor isoforms. *FEBS Lett.* 278 239-243.
- MOUSLI, M., BUEB, J., BRONNER, C., ROUOT, B., and LANDRY, Y. (1990). G protein activation: A receptor-independent mode of action for cationic amphiphilic neuropeptides and venom peptides. *Trends Pharmacol. Sci.* 11, 358-362.
- MOYLE, W.R., BERNARD, M.P., MYERS, R.V., MARKO, O.M., and STRADER, C.D. (1991). Leutropin/beta-adrenergic receptor chimeras bind choriogonadotropin and adrenergic ligands but are not expressed at the cell surface. *J. Biol. Chem.* 266, 10807-10812.
- MUNCH, G., DEES, C., HEKMAN, M., and PALM, D. (1991). Multisite contacts involved in coupling of the β -adrenergic receptor with the stimulatory guanine-nucleotide-binding regulatory protein. *Eur. J. Biochem.* 198, 357-364.
- MURPHY, P.M., and TIFFANY, H.L. (1991). Cloning of complementary DNA encoding a functional human interleukin-8 receptor. *Science* 253, 1280-1283.
- NAGAYAMA, Y., RUSSO, D., CHAZENBALK, G.D., and RAPOPORT, B. (1991a). Seven amino acids (lys-201-lys-211) and 9 amino acids (gly-222 to leu-230) in the human thyrotropin receptor are involved in ligand binding. *J. Biol. Chem.* 266, 14926-14930.
- NAGAYAMA, Y., WADSWORTH, H.L., CHAZENBALK, G.D., RUSSO, D., SETO, P., and RAPOPORT, B. (1991b). Thyrotropin-luteinizing hormone/chorionic gonadotropin receptor extracellular domain chimeras as probes for thyrotropin receptor function. *Proc. Natl. Acad. Sci. USA* 88, 902-905.
- NATHANS, J., and HOGNESS, D.S. (1984). Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proc. Natl. Acad. Sci. USA* 81, 4851-4855.
- NATHANS, J., THOMAS, D., and HOGNESS, D.S. (1986). Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science*, 232, 193-202.
- NEVE, K.A. (1991). Regulation of dopamine D2 receptors by sodium and pH. *Mol. Pharmacol.* 39, 570-478.
- NEVE, K.A., COX, B.A., HENNINGSEN, R.A., SPANOYANIS, A., and NEVE, R.L. (1991). Pivotal role for aspartate-80 in the regulation of dopamine D2 receptor affinity for drugs and inhibition of adenylyl cyclase. *Mol. Pharmacol.* 39, 733-739.
- NISHIMOTO, I., MURAYAMA, Y., KATADA, T., UI, M., and UGATA, E. (1989). Possible direct linkage of insulin-like growth factor-II receptor with guanine nucleotide binding proteins. *J. Biol. Chem.* 264, 14029-14038.
- NISHIMOTO, I., OGATA, E., and OKAMOTO, T. (1991). Guanine nucleotide-binding protein interacting but unstimulating sequence located in insulin-like growth factor II receptor. *J. Biol. Chem.* 266, 12747-12751.
- O'DOWD, B.F., HNATOWICH, M., REGAN, J.W., LEADER, W.M., CARON, M.G., and LEFKOWITZ, R.J. (1988). Site-directed mutagenesis of the cytoplasmic domains of the human β_2 -adrenergic receptor. *J. Biol. Chem.* 263, 15985-15992.
- O'DOWD, B.F., HNATOWICH, M., CARON, M.G., LEFKOWITZ, R.J., and BOUVIER, M. (1989a). Palmitoylation of the human β_2 -adrenergic receptor. *J. Biol. Chem.* 264, 7564-7569.
- O'DOWD, B.F., LEFKOWITZ, R.J., and CARON, M.G. (1989b). Structure of the adrenergic and related receptors. *Annu. Rev. Neurosci.* 12, 67-83.
- OHNO, S. (1970). *Evolution by gene duplication*. (Springer, New York).
- OKAMOTO, T., KATADA, T., MURAYAMA, Y., UI, M., OGATA, E., and NISHIMOTO, I. (1990). A simple structure encodes G protein-activating function of the IGF-II/mannose 6-phosphate receptor. *Cell* 62, 709-717.
- OVCHINNIKOV, Y.A., ABDULAEH, N.G., FEIGINA, M.Y., ARTAMONOV, I.D., ZOLOTEREVE, A.S., KOSTINA, M.B., BOGACHUK, A.S., MIROSHNIKOV, A.I., MARTI-

- NOV, V.I., and KUDELIN, A.B. (1982). The complete amino acid sequence of visual rhodopsin. *Bioorg. Khim.* 8, 1011-1014.
- OVCHINNIKOV, Y.A., ABDULAEV, N.G., and BOGACHUK, A.S. (1988). Two adjacent cysteine residues in the C-terminal cytoplasmic fragment of bovine rhodopsin are palmitylated. *FEBS LETT.* 230, 1-5.
- PALM, D., MUNCH, G., DEES, C., and HEKMAN, M. (1989). Mapping of β -adrenoceptor coupling domains to G_s -protein by site-specific synthetic peptides. *FEBS Lett.* 254, 89-93.
- PALM, D., MUNCH, G., MALEK, D., DEES, C., and HEKMAN, M. (1990). Identification of a G_s -protein coupling domain to the β -adrenoceptor using site-specific synthetic peptides: Carboxylterminus of the G_{sa} is involved in coupling to β -adrenoceptors. *FEBS Lett.* 261, 294-298.
- PERALTA, E.G., ASHKENAZI, A., WINSLOW, J.W., SMITH, D.H., RAMACHANDRAN, J., and CAPON, D.J. (1987). Distinct primary structures, ligand-binding and tissue expression of four human muscarinic acetylcholine receptors. *EMBO J.* 6, 3923-3929.
- REGAN, J.W. (1988). In *The α_2 -Adrenergic Receptors*. L.E. Limbird, ed. (Humana, Clifton, NJ). pp. 15-74.
- REGAN, J.W., KOBILKA, T.S., YANG, F.T.L., CARON, M.G., and KOBILKA, B.K. (1988). Cloning and expression of a human kidney cDNA for an alpha 2-adrenergic receptor subtype. *Proc. Natl. Acad. Sci. USA* 85, 6301-6305.
- ROSS, P.C., FIGLER, R.A., CORJAY, M.H., BARBER, C.M., ADAM, N., HARCUS, D.R., and LYNCH, K.R. (1990). RTA, a candidate G protein-coupled receptor: Cloning, sequencing, tissue distribution. *Proc. Natl. Acad. Sci. USA* 87, 3052-3056.
- RUSSO, D., CHAZENBALK, G.D., NAGAYAMA, Y., WADSWORTH, H.L., and RAPOPORT, B. (1991). Site-directed mutagenesis of the human thyrotropin receptor: Role of asparagine-linked oligosaccharides in the expression of a functional receptor. *Mol. Endocrinol.* 5, 29-33.
- SAKMAR, T.R., FRANKE, R.R., and KHORANA, H.G. (1989). Glutamic acid-113 serves as the retinylidene Schiff base counterion in bovine rhodopsin. *Proc. Natl. Acad. Sci. USA* 86, 8309-8313.
- SAKURAI, T., YANAGISAWA, M., TAKUWA, Y., MIYAZAKI, H., KIMURA, S., GOTO, K., and MASAKI, T. (1990). Cloning of a cDNA encoding a non-isopeptide-selective subtype of the endothelin receptor. *Nature* 348, 732-735.
- SASAKI, K., YAMANO, Y., BARDHAN, S., IWAI, N., MURRAY, J.J., HASEGAWA, M., MATSUDA, Y., and INAGAMI, T. (1991). Cloning and expression of a complementary DNA encoding a bovine adrenal angiotensin II type-1 receptor. *Nature* 351, 230-233.
- SCHWINN, D.A., LOMASNEY, J.W., LORENZ, W., SZKLUT, P.J., YANG, F.T.L., CARON, M.G., LEFKOWITZ, R.J., and COTECCHIA, S. (1990). Molecular cloning and expression of the cDNA for a novel 1-adrenergic receptor subtype. *J. Biol. Chem.* 265, 8183-8189.
- SHAPIRO, R.A., WAKIMOTO, B.T., SUBERS, E.M., and NATHANSON, N.M. (1989). Characterization and functional expression in mammalian cells of genomic and cDNA clones encoding a *Drosophila* muscarinic acetylcholine receptor. *Proc. Natl. Acad. Sci. USA* 86, 9039-9043.
- SHIGEMOTO, R., YOKOTA, Y., TSUCHIDA, K., and NAKANISHI, S. (1990). Cloning and expression of a rat neuromedin K receptor cDNA. *J. Biol. Chem.* 265, 623-628.
- SNYDER, L.A., ROBERTS, J.L., and SEALFON, S.C. (1991). Alternative transcripts of the rat and human dopamine D3 receptor. *Biochem. Biophys. Res. Commun.* 180, 1031-1035.
- SOKOLOFF, P., GIROS, B., MARTRES, M.P., BOUTHENET, M.L., and SCHWARTZ, J.C. (1990). Molecular cloning and characterization of a novel dopamine receptor (D3) as a target for neuroleptics. *Nature* 347, 146-151.
- SPINDEL, E.R., GILADI, E., BREHM, P., GOODMAN, R.H., and SEGERSON, T.P. (1990). Cloning and functional characterization of a complementary DNA encoding the murine fibroblast bombesin/gastrin-releasing peptide receptor. *Mol. Endocrinol.* 4, 1956-1963.
- SPRENGEL, R., BRAUN, T., NIKOLICS, K., SEGALOFF, D.L., and SEEBURG, P.H. (1990). The testicular receptor for follicle stimulating hormone: Structure and functional expression of cloned cDNA. *Mol. Endocrinol.* 4, 525-530.
- SREEDHARAN, S.P., ROBICHON, A., PETERSON, K.E., and GOETZL, E.J. (1991). Cloning and expression of the human vasoactive intestinal peptide receptor. *Proc. Natl. Acad. Sci. USA* 88, 4986-4990.
- STRADER, C.D., DIXON, R.A.F., CHEUNG, A.H., CANDELORE, M.R., BLAKE, A.D., and SIGAL, I.S. (1987a). Mutations that uncouple the β -adrenergic receptor from G_i and increase agonist affinity. *J. Biol. Chem.* 262, 16439-16443.
- STRADER, C.D., SIGAL, I.S., REGISTER, R.B., CANDELORE, M.R., and RANDS, D.R.A. (1987b). Identification of residues required for ligand binding to the beta-adrenergic receptor. *Proc. Natl. Acad. Sci. USA* 84, 4384-4388.
- STRADER, C.D., SIGAL, I.S., CANDELORE, M.R., RANDS, E., HILL, W.S., and DIXON, R.A.F. (1988). Conserved aspartate residues 79 and 113 of the β -adrenergic receptor have different roles in receptor function. *J. Biol. Chem.* 263, 10267-10271.
- STRADER, C.D., CANDELORE, M.R., HILL, W.S., SIGAL, I.S., and DIXON, R.A.F. (1989a). Identification of two serine residues involved in agonist activation of the β -adrenergic receptor. *J. Biol. Chem.* 264, 13572-13578.
- STRADER, C.D., SIGAL, I.S., and DIXON, R.A.F. (1989b). Structural basis of β -adrenergic receptor function. *FASEB J.* 3, 1825-1832.
- STRASSER, R.H., STILES, G.L., and LEFKOWITZ, R.J. (1984). Differences in the carbohydrate chains of mammalian β_1 - and β_2 -adrenergic receptors. *Circulation* 70, 68.
- STRAUB, R.E., FRECH, G.C., JOHO, R.H., and GERSCHEGORN, M.C. (1990). Expression cloning of a cDNA encoding the mouse pituitary thyrotropin-releasing hormone receptor. *Proc. Natl. Acad. Sci. USA* 87, 9514-9518.
- STRITTMATER, S.M., VALENZUELA, D., KENNEDY, T.E., NEER, E.J., and FISHMAN, M.C. (1990). G_o is a major growth cone protein subject to regulation by GAP-43. *Nature* 344, 836-841.
- SUNAHARA, R.K., GUAN, H.C., O'DOWD, B.F., SEEMAN, P., LAURIER, L.G., NG, G., GEORGE, S.R., TORCHIA, J., VAN TOL, H.H., and NIZNIK, H.B. (1991). Cloning of the gene for a human dopamine D5 receptor with higher affinity for dopamine than D1. *Nature* 350, 614-619.
- TAKAHASHI, N., TAKAHASHI, Y., and PUTNAM, F.W. (1985). Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich α_2 -glycoprotein of human serum. *Proc. Natl. Acad. Sci. USA* 82, 1906-1910.
- TAKASU, N., CHARRIER, B., MAUCHAMP, J., and LISSITZKY, S. (1978). Modulation of adenylate cyclase/cAMP response to thyrotropin and prostaglandin E2 in cultured thyroid cells: Positive regulation. *Eur. J. Biochem.* 90, 139-146.
- TANAKA, K., MASU, M., and NAKANISHI, S. (1990). Structure and functional expression of the cloned rat neurotensin receptor. *Neuron* 4, 847-854.
- THOMAS, D.D., and STRYER, L. (1982). Transverse location

- of the retinal chromophore of rhodopsin in rod outer segment disc membranes. *J. Mol. Biol.* 154, 145-157.
- TOTA, M.R., and STRADER, C.D. (1990). Characterization of the binding domain of the β -adrenergic receptor with the fluorescent antagonist carazolol. *J. Biol. Chem.* 265, 16891-16897.
- TSAI-MORRIS, C.H., BUCZKO, E., WAND, W., and DUFAU, M.L. (1990). Intronic nature of the rat luteinizing hormone receptor gene defines a soluble receptor subspecies with hormone binding activity. *J. Biol. Chem.* 265, 19385-19388.
- UNWIN, P.N., and HENDERSON, R. (1975). Molecular structure determination by electron microscopy of crystalline specimens. *J. Mol. Biol.* 94, 425-440.
- VAN-TROL, H.H., BUNZOW, J.R., GUAN, H.C., SUNAHARA, R.K., SEEMAN, P., NIZNIK, H.B., and CIVELLI, O. (1991). Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine. *Nature* 350, 610-614.
- VENTER, J.C., FRASER, C.M., KERLAVAGE, A.R., and BUCK, M.A. (1989). Molecular biology of adrenergic and muscarinic cholinergic receptors. *Biochem. Pharmacol.* 38, 1197-1208.
- VOIGT, M.M., KISPERS, J., and CHIN, H.M. (1990). Sequence of a rat brain cDNA encoding an alpha-1B receptor. *Nucleic Acids Res.* 18, 1053.
- VU, T.-H.H., HUNG, D.T., WHEATON, V.I., and COUGHLIN, S.R. (1991). Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* 64, 1057-1068.
- WADA, E., WAY, J., SHAPIRA, H., KUSANO, K., LEBACQZ, V.A.M., COY, D., JENSEN, R., and BATTERY, J. (1991). cDNA cloning, characterization, and brain region-specific expression of a neuromedin-B-preferring bombesin receptor. *Neuron* 6, 421-430.
- WADSWORTH, H.L., CHAZENBALK, G.D., NAGAYAMA, Y., and RUSSO, D. (1990). An insertion in the human thyrotropin receptor critical for high affinity hormone binding. *Science* 249, 1423-1425.
- WALKER, R.J., and HOLDEN-DYE, L. (1989). Commentary on the evolution of transmitters, receptors and ion channels in invertebrates. *Comp. Biochem. Physiol.* 93A, 25-39.
- WANG, C.D., BUCK, M.A., and FRASER, C.M. (1991). Site-directed mutagenesis of alpha2A-adrenergic receptors: Identification of amino acids involved in ligand binding and receptor activation by agonists. *Mol. Pharmacol.* 40, 168-179.
- YOKOTA, Y., SASAI, Y., TANAKA, K., FUJIWARA, T., TSUCHIDA, K., SHIGEMOTO, R., KAKIZUKA, A., OHKUBO, H., and NAKANISHI, S. (1989). Molecular characterization of a functional cDNA for rat substance receptor. *J. Biol. Chem.* 264, 17649-17652.
- YOKOYAMA, S., ISENBERG, K.E., and WRIGHT, A.F. (1989). Adaptive evolution of G-protein coupled receptor genes. *Mol. Biol. Evol.* 6, 342-353.
- YOKOYAMA, R., and YOKOYAMA, S. (1990). Convergent evolution of the red- and green-like visual pigment genes in fish, *Astyanax fasciatus*, and human. *Proc. Natl. Acad. Sci. USA* 87, 9315-9318.
- YOUNG, D., WAITCHES, G., BIRCHMEIER, C., FASANO, O., and WIGLER, M. (1986). Isolation and characterization of a new cellular oncogene encoding protein with multiple potential transmembrane domains. *Cell* 45, 711-719.
- ZHUKOVSKY, E.A., and OPRIAN, D.D. (1989). Effect of carboxylic acid side chains on the absorption maximum of visual pigments. *Science* 246, 928-930.

Address reprint requests to:

Dr. Stuart C. Sealfon
Neurobiology-Box 1065
Mount Sinai School of Medicine
One Gustave L. Levy Place
New York, NY 10029

Received for publication December 13, 1990, and in revised form October 23, 1991.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.